

Where UK academics go next

British universities will soon be at another turning point. The danger is that they will succumb to financial pressure; but they have an opportunity to get a grip on their own affairs.

THE continuing agony of the British university system, fast becoming a bore for others, will be advanced a further notch in the next few weeks. By now, the chairman of the University Grants Committee, Sir Peter Swinnerton-Dyer, is well embarked on his self-imposed task of reading the replies that he has received to the questionnaire put out in November to universities and anybody else who cared to answer it. No doubt the grants committee will eventually devise a policy for the years ahead, perhaps even a policy that will be understood by the universities and the British Government as a set of principles around which their future relationship will be organized. Here is the least of what the universities should be expecting.

Security. The British Government has been playing cat and mouse with university budgets. Fees for overseas students were by external fiat required to be increased to what the government considered an "economic" level in 1980, but the predicted effect of falling numbers was offset two years later by a £40 million fund in the gift of the Foreign and Commonwealth Office. A few months later, a progressive reduction over three years of 8.5 per cent in university budgets was decreed (with effect from the academic year beginning in 1981), but partly offset eighteen months ago by extra funds for appointing young people to academic staffs. During this period the government seemed to promise stability, but by a year ago was canvassing the notion that there might have to be a further erosion of university support either on demographic grounds (the student age-group will be declining) or even just for its own sake. These possibilities, which appear as hypotheses in the grants committee's questionnaire, have been widely and rightly disputed, notably by the Royal Society (see p.488). So they should be, in a setting in which the shortage of skill is overwhelming. But the least that the universities should demand is to know where they stand more than a few months in advance. (Some establishments, notably the components of the University of London, have still not been told their subventions for the financial year beginning in August, but have only until September to calculate how many academics must be forced into early retirement to take advantage of the financial arrangements which then lapse.) The idea that it is possible to run academic institutions as if they were retail shops is intolerable, unworkable and wasteful.

Freedom. The worst feature of the system forced on British universities in the past three years is that each of them has been given (by the grants committee) a quota of students which, if exceeded, will invite financial penalties. The original objective was twofold — to give individual universities some idea of the scale on which they would be helped financially to operate, and to assist the government in controlling the otherwise unquantifiable bill the government pays through local authorities for student maintenance. While the issue is not mentioned in the famous questionnaire, the consequences are uneconomic (educational plant is partly unused), inequitable (British students who could pay full costs, and who might benefit, are kept out) and an intrusion on academic freedom that should worry the universities more than it seems to have done. The present government is unlikely to budge in its determination to peg the total cost of higher education, but is probably indifferent to the way in which money is spent. The universities, devoted as they are to the principle of open access for all qualified students and fearful of schemes for student loans, have been slow to devise alternatives. Now the vice-chancellors

have set up a committee to consider alternative ways of financing higher education. The outcome could be the chance to put British universities on a sound basis.

Research. The vitality of university research has been the chief casualty of the past four years, and one of the grants committee's questions asks for academic opinions on the subject. Most universities have replied that the system for supporting research by two channels — through university budgets and through project grants — has broken down because funds intended for research support are being spent on other things. So much is true. But the largest single element in grants committee support for research is notional, the fraction of academic salaries corresponding to the time that university teachers supposedly spend on research. Two years ago, when a committee under Sir Alec Merrison acknowledged that the system had broken down, it urged that universities should create mechanisms for deciding how the available resources for research should be allocated, but very little has yet been done. Until institutions have internal mechanisms commanding mutual respect for explicit allocations of effort, they will not be equipped to plot a coherent strategy (or to spend the larger funds that they deserve).

Diversity. It is no longer feasible for the British university system to pretend that all universities are alike. The unremarkable reality, apparent for years, is that they differ from each other in a host of different ways, but especially in their flair for research. In tacit acknowledgement of this, the annual grants doled out to British universities have hitherto included an ostentatiously undefined element dealing more generously with research universities than with others. The grants committee has now asked whether this practice should be made explicit — whether some part of the annual budget should be "earmarked" as the saying goes. The Royal Society's opinion on the subject (see p.488), like that of most universities, is that earmarking would be an unacceptable infringement of freedom (as it would be if institutions were instructed what to teach, or how). The vice-chancellors are more openly opposed, on the grounds that teaching and research are indivisible. But the choice is between covert (as now) and explicit earmarking. On the principle that institutions will be best served if they know where they stand, earmarking is now obviously preferable. What university freedom demands is that it should be possible for universities even within a publicly financed system such as the British to work their way up the ladder, which argues for a mechanism based on past success in attracting project grants, and for a modest transfer of project-grant funds (say 10 per cent) from research councils to the university system (so that the research committees the universities should be setting up can have something in the kitty).

Flexibility. In the new circumstances, universities need above all to be able to make their way in the world without artificial restraints. This argues for the removal of a host of existing barriers — the division of the system of higher education into two, nationally negotiated salary scales for academics and the others who work in the universities, unreasonable external restraints on the conversion of existing capital assets into new income and the like. One simplifying course of action would be to rename as universities the polytechnics that make up the other half of the British higher education system. (Some of them might, of course, prefer to stay the way they are.) That would require decision by



the government, now reluctantly up to its eyes in the administration of these institutions. But the universities have it entirely within their own gift to work out for themselves, and with their employees, more flexible ways of working. Even before the University Grants Committee has finished reading what people have been telling it, that is what the universities should be doing. □

No common market

Europe seems to have learned very little from the traumatic experience of recent weeks.

CAPITALISTS are supposedly taught at Midas's knee about the laws of supply and demand, and Brussels is one of the minor capitals of European capitalism, which is why it is incongruous that European ministers of agriculture, meeting there at the weekend, should have gone out of their way to fly in the face of economic reason. After the failure of the meeting of heads of government of the European Communities (see *Nature* 15 March, p.213), a meeting of the agriculture ministers was made necessary by the need to cut the cost of agricultural subsidies to stay roughly within the budget likely to be available.

The problem is by now well known. The European Commission, as buyer of the last resort, is compelled to buy food from farmers at agreed prices, which is why it has been accumulating huge amounts of surplus food. For the past several years, the most conspicuous surplus has been of milk and materials made from it — Europe now has a stockpile of nearly a million tonnes of butter for example. So innocent readers of economic textbooks would have expected the meeting at the weekend to have reduced the support price for milk, reducing the supply (by decreasing the incentive for farmers to produce surplus milk) and increasing the demand (by making milk products cheaper for consumers). But in the event, the ministers of agriculture settled for a different scheme. While the prices of most foods were marginally reduced, the embarrassment about milk is primarily to be dealt with by means of production quotas. If farmers in member states collectively produce more than intended, Brussels will no longer take the surplus off the market. Somehow, national governments will have to penalize over-productive farmers, but even so it seems that, taken all together, the quotas for next year amount to 10 million tonnes more milk than Europe can possibly consume. One predictable result is that milk-producing farmers are up in arms. Another is that there remains serious doubt whether the European Commission will be able to meet the cost of the policy now agreed. But there is no plan that European consumers, the people who might be able to strike some balance between supply and demand by increasing their consumption, are to be given an incentive to do so. Instead, in due course, they will be expected to pay more in taxes to buy up the milk surplus still in prospect. This is the economics of the looking-glass.

Why in these circumstances has the agreement at the weekend been widely welcomed at the beginning of the long process for reforming European policy on agriculture? Those now punch-drunk from too many meetings in Brussels are saying that, for the first time in thirty years, it has been acknowledged that the scale of European agricultural production should not be determined solely by the capacity of farmers to produce at prices which are adjusted from year to year by the inflation rate. For the first time, these weary optimists are saying, it has been acknowledged that the capacity to consume should play a part in the equation. In due course, the argument goes, it may even be possible to arrange that demand should influence supply through the prices that people are prepared to pay for commodities such as milk. But if in due course, why not now? Nobody asks that farmers should be so harshly dealt with that they are forced out of business suddenly, on a change of whim in Brussels, but if this is the beginning of a realistic market in agricultural produce, it is surely better that farmers should be given signals from the ultimate consumers and not from the bureaucrats by whom they are represented.

Those concerned will say their task would have been simpler if they had had more time, but in reality there have been thirty years

(since the beginning of the European Economic Community) and certainly thirty months (since it became plain that Europe would be in a crisis this year). And agriculture is only one of several outstanding issues that must be settled either at the next meeting of the heads of government in Brussels in June or at some earlier all-night session. Is it too late to ask that the governments concerned should broaden the issues they are discussing, perhaps by linking agriculture with bothersome social and economic problems, in the hope that it will then be easier to make the compromises that are necessary and, at the same time, to reach understanding on the principles of their common activities? Agricultural policy in Europe, for example, should be a way of protecting producers against the short-term fluctuation over, say three to five years, not long-term schemes for subsidizing them. And there is just as urgent a need that social problems such as unemployment should be jointly mitigated but within similar time-scales. □

Tokenism does no good

Cosmetic devices for making engineering more prominent in US research should be abandoned.

LIKE many previous advocates of minority rights, Mr George Brown, the California Democrat who is forever urging that the US House of Representatives should do something to rescue engineering and technology from neglect, has been driven back on tokenism. At various times in the past six years, Mr Brown has sought to breathe life into a project for setting up a federal agency, analogous to the National Science Foundation, that would support research in engineering and technology. The public hearings on the subject over the years have served to demonstrate that there are indeed serious problems to be tackled in the organization of engineering education at United States universities, but have failed to show that a new grant-making agency would cure them. Now, Mr Brown seems to have fallen back on an entirely meaningless gesture: renaming the National Science Foundation so as to emphasize its responsibilities for engineering or technology (see p.486). It is not surprising that he should have been defeated.

Mr Brown should first have looked more carefully at what has been happening in the United Kingdom, where there used to be an organization called the Science Research Council with responsibilities for supporting research in science and engineering. Three years ago, under pressure from the UK Conference of Engineering Professors and the then newly created Fellowship of Engineering, both of which were crying for an Engineering Research Council, the Science Research Council changed its name, adding the word engineering in the middle. The consequences of this strictly cosmetic device have been negligible. Spending by the Science Research Council on the support of engineering research in British universities had been increasing year by year for nearly a decade before the change of name was considered expedient. Now, with its budget under pressure, the renamed council has rightly decided (see *Nature* 29 March, p.391) to call a halt to this growth.

Changing the name of the National Science Foundation would probably be similarly pointless. The problems that most bother engineering schools in the United States — the difficulties of recruiting and keeping able faculty members in the face of competition from industry and of administering cooperative projects with industry in such a way that they are educationally as well as commercially worthwhile — will not be solved by a grant-making agency, however generously endowed. Moreover, there is plenty of cautionary experience in the recent history of the National Science Foundation to show that artificially imposed patterns of research spending are often unworkable. The foundation's own artificial programme launched in the 1970s was a predictable disaster, long since discreetly abandoned. And while it is true that merely changing the foundation's name should be less directly harmful and, indeed, should not necessarily have any influence at all, Mr Brown and his friends should now let the proposal drop. □

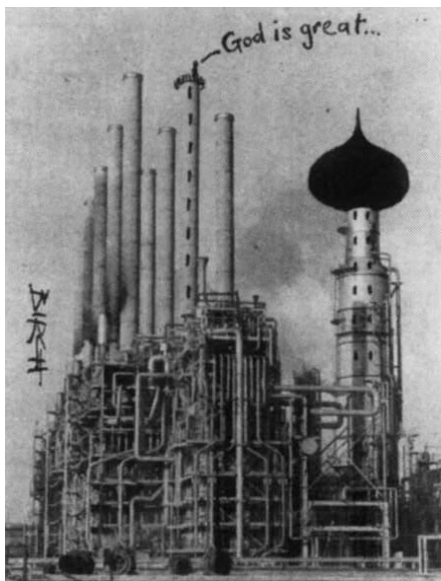
Chemical weapons

United Nations accuses Iraq of military use

Washington

EVIDENCE of Iraq's use of chemical weapons has been produced by a team of experts sent to Iran by the United Nations (UN). In a report issued last week, the team said that it had examined a number of unexploded or damaged bombs in the war zone in Iran, all of which were designed to carry liquid, and that samples collected from one of the bombs proved upon analysis in Swedish and Swiss laboratories to be mustard gas. Soil and liquid samples provided to the team by Iranian forces and said to have come from a leaking bomb at another site in the war zone were found to contain the nerve gas tabun, also known as GA.

The UN team also examined Iranian soldiers at four hospitals in Iran; all 35 in the first group examined showed symp-



toms that the team said could be caused only by vesicant agents such as mustard gas. These symptoms included large areas of reddened and blistered skin, conjunctivitis and low white blood cell count. A second group of six patients was examined following the reported attack from which the GA sample was obtained; this group showed signs of nerve-gas poisoning, including respiratory problems, vomiting, tremors and acute miosis (contraction of the pupils). According to analyses run by the Iranian hospital staff, two of the six had very low acetylcholinesterase levels in their blood.

The US State Department says that it has known since late last year that Iraq has been manufacturing and using chemical weapons. It says that after quiet persuasion failed, it made its charges public in early March. A State Department official said

last week that besides GA and mustard gas, Iraq is also now attempting to produce GB, or sarin, a nerve agent which is considerably more lethal than either of the other two. (The US Army considers both GA and mustard gas to be obsolete but does stock GB).

Mustard gas is easily manufactured in one step from thiodiglycol and hydrochloric acid. GA is also relatively easy to produce; it requires a two-step synthesis using the materials ethanol, dimethylamine, potassium cyanide and phosphorous oxychloride. Sophisticated facilities, such as those used in advanced pesticide plants, are however needed to protect the workers in a plant producing GA. All of the nerve agents are chemically related to organophosphorous pesticides.

The bomb casings used in Iraq to deliver the chemical agents appear to be of a standard US pattern. The UN team reported that all the bombs it examined were identical; green, with a 10-cm yellow band around the nose, and the marking "BR 250 WP". The team said that the thinness of the bomb casings — 1 to 2 mm — indicated that they were not intended for use as high-explosive bombs.

According to Tom Gervasi, an expert on the arms trade who runs the Center for Military Research and Analysis in New York, the markings and dimensions of the bomb correspond to a standard 250-pound US bomb used for white phosphorus — hence the marking "WP". Gervasi noted that many other NATO countries also manufacture the bomb casings. White phosphorus, an incendiary, is normally

packed in thin-walled casings; the casing is effective for dispersing chemical agents as well.

The mustard gas samples were also analysed for mycotoxins — the fungal toxins that the United States claims have been used by Soviet and Soviet-backed forces in Afghanistan and South-East Asia — but none was detected. Press reports in recent weeks have claimed that Iraq was using these "yellow rain" toxins as well as chemicals. Dr Aubin Heyndrickx of the Ghent Toxicological Institute, Belgium, added credence to these reports by announcing that he had detected mycotoxins in the tenth of a part per million (p.p.m.) range in blood samples drawn from Iranian soldiers being treated in Vienna.

Heyndrickx's claims have, however, been greeted by other scientists with considerable scepticism — scepticism that has been fuelled by Heyndrickx's assertion that scientists at Sweden's National Defence Research Institute have confirmed his results. In fact, the Swedish scientists vigorously deny that they have been able to detect any mycotoxins in the samples they have looked at. The UN findings on mycotoxins were based on thin-layer chromatography, which has a detection limit of 5 p.p.m.

Although the UN report seems to leave little doubt that chemical weapons have been used, it is careful not to draw any conclusions about who is responsible. And the possibility, however remote, of fabrication of evidence by Iran has not yet been completely ruled out.

The UN team was made up of Dr Gustav Andersson of Sweden's National Defence Research Institute, Dr Manuel Dominguez of Spain's Army Medical Corps, Dr Peter Dunn of Australia's Defence Department Materials Research Laboratories, and Colonel Ulrich Imobersteg, chief of Switzerland's chemical defence division.

Stephen Budiansky

Arms control advisers wanted

Los Angeles

THE Carnegie Corporation of New York has just awarded Stanford University's Center for International Security and Arms Control a four-year grant to train two or three young scientists a year in the art of giving advice to government on arms control. Dr Sidney Drell, a co-director of the centre and deputy director of the Stanford Linear Accelerator, will run the project.

The objective is to fill the gap left by the ageing of the small group of scientists who, since the Manhattan Project, have been dispensing advice to presidents, the military and Congress on how to control the arms race. "We have not in my judgement been getting enough young scientists involved in arms control", Dr Drell says.

Like a handful of others from his generation, Dr Drell, now 57, was first invited to

advise the government on arms control in the early 1960s. But more recently, he says, fewer young scientists have been given a chance to apply scientific knowledge to security issues. And while the huge cadre of defence scientists working directly for the government do excellent work, Drell says, the need that independent scientists should be working on defence issues is more important than ever.

Drell thinks others among the "older generation" of experts — people like Hans Bethe, Jerome Wiesner, Paul Doty, Frank Long and Harvey Brooks — will also seek out younger scientists to train in the art of arms control. Leading centres for such work are at Harvard University, Massachusetts Institute of Technology, Cornell University, the University of California at Los Angeles and Stanford University.

Sandra Blakeslee

Research reactors

NRC rule may backfire

Washington

THE only consequence of a worthy attempt by the Nuclear Regulatory Commission (NRC) to promote the cause of nuclear non-proliferation may be to damage research and training in universities. That is the fear of many of the 25 universities in the United States which operate research reactors and which expect to be asked, in a draft NRC rule this month, to convert them from high to low enriched uranium fuel.

The new rule is ostensibly intended to reduce the risks that terrorists might steal enough weapons-grade uranium to make an atomic bomb, but the universities believe the threat of terrorism is already minimal and that the NRC action is politically motivated.

The 25 reactors are widely used. One of the largest, the 10-MW reactor operated by the Massachusetts Institute of Technology (MIT), generates several million dollars a year of sponsored research for the institute. In addition to training nuclear engineers, MIT uses the facility for neutron analysis of trace elements, for the production of medical isotopes and for research on power reactor safety.

Dr Otto Harling, director of MIT's nuclear reactor laboratory, said last week that the danger of terrorism had been exaggerated. He claimed that none of the university reactors was allowed to store more than 5 kilograms of unirradiated fuel, and that at least 15 kilograms of uranium was necessary to produce a crude nuclear weapon. To steal enough uranium for a bomb, terrorists would have to attack several reactors at once or — a complex and hazardous undertaking — remove burned fuel from the reactor core.

Although the Department of Energy (DoE) is spending some \$5 million a year on development of new kinds of low-enriched uranium for research reactors, it appears to agree with the universities' assessment. Testifying in the Senate last month, Dr Alvin Trivelpiece, director of research at DoE, said he did not believe there was much risk of sabotage or theft at the university facilities. NRC, however, is expected to press ahead with the rule as a gesture designed to impress foreign countries with the United States' concern about the dangers of nuclear proliferation.

Many of the 25 university reactors use uranium that is more than 90 per cent enriched. A university study group set up by NRC last year concluded that several of them could be converted to use a new 20 per cent enriched fuel developed by the Argonne National Laboratory, but Mr Lincoln Clark, director of reactor operations at MIT and a member of the panel, said the universities doubted whether the exercise would be useful, since universities accounted for only 10 per cent of the

weapons-grade uranium used in research reactors. The bulk, used by DoE's research reactors, would not be affected by the rule.

The universities also fear that the measure could backfire. One worry is the estimated \$15 million cost of conversion. Unless DoE foots the bill, several financially hard-pressed universities are likely to close their reactors rather than convert them. And according to Dr Harling, it is all too probable that anti-nuclear groups would take advantage of the public hearings that would be required when relicensing converted reactors to try to close university reactors altogether.

NRC has told the universities that it will try to make any relicensing hearings as smooth as possible, but the universities doubt whether any short cuts are legally possible. Meanwhile, there are knotty technical obstacles to be overcome. Several of the reactors were designed with a lifetime fuel core, and conversion will therefore be difficult. And the two biggest reactors — at MIT and Missouri — cannot be modified because there is not enough room in their cores to accommodate the larger volume of low-enriched uranium.

Peter David

IAEA steps in

THE International Atomic Energy Agency is to prepare a safety guide laying down principals for decommissioning research reactors. The agency points out that 100 of the 270 research and test reactors in operation worldwide are more than 20 years old and so approaching the end of their useful lives. Although there are no special difficulties in decommissioning research reactors that have not been met in power reactors, experience so far is limited to industrialised countries. And the reactors are very varied in their construction.

Most of the new research reactors being built are in developing countries, and the agency wants clear criteria for decommissioning to be established and incorporated into reactor designs. In both Britain and the United States the shrinking nuclear power industries have resulted in a lower demand for nuclear engineers, and financial pressures on higher education institutions have made reactors seem very expensive training tools.

A teaching reactor at Queen Mary's College in London was decommissioned a few years ago because the college could not afford to run it. Others are facing a 15 per cent down-grading of their thermal flux ratings because of the need to convert to low enriched fuel; for some applications this could mean a real loss of capability. The pressure to change to low enriched fuel is blamed on US restrictions on the export of very highly enriched fuel. Tim Beardsley

Aachen hospital

High hopes — higher costs

THE Gross-Klinikum of the Technische Hochschule at Aachen in West Germany (see *Nature* 297, 267; 1982) is more bedevilled by financial problems than ever before in its 15-year gestation. Estimated completion costs have risen to more than DM 2,400 million (£635 million) and on the advice of a still secret report of the *Bundesrechnungshof*, the West German Ministry of Education and Science is refusing to pay more than its 50 per cent share of a previously agreed limit of DM 1,700 million. Eventual running costs of DM 500–600 million a year will be borne entirely by the *Land*.

The eight-storey building — the so-called *Bettenberg* — covers the area of five football pitches and contains 52 operating theatres and 1,500 beds. It will be devoted 44 per cent to research and teaching and 66 per cent to hospital functions, replacing the ageing medical facilities of the border city of Aachen. It is unique in Europe in providing integrated health care, research and educational facilities under one roof.

Medical focus will be particularly on kidney disease, heart disease, paediatric cardiology, burns and plastic surgery. The clinic will have a yearly intake of 450 medical students, 80 dental students and 500 students in ancillary professions.

Clinical research, will emphasize biomedical technology, driving mechanisms for artificial hearts, synthetic valves, vibration techniques for removal of kidney stones and neuroradiology.

Conceived by the *Wissenschaftsrat* in the 1960s in a mistaken estimate of the health care needs of the Aachen area, and on an erroneous assumption that it would draw patients from outside Germany, many of the massive costs can be laid at the door of the concept of *Synchronplanung*, meaning planning while you build. Inflation, changes in building regulations and the needs of progress in medical and pharmaceutical knowledge, added to a large measure of error, omission and general incompetence and mechanical subsidence, have produced an unparalleled financial situation. There is litigation between the *Land* and the builders *Neue Heimat Stadtebau*, the *Land* does not accept the report to the federal ministry, which is refusing to raise its contribution partly on the grounds that this could curtail its contribution to other institutions in North Rhine-Westphalia, and there are fears that, having siphoned off 230 academic jobs from universities in the *Land*, the clinic may have a continued unhealthy influence. The general secretary of the Free Democratic Party, Frau Adam-Schwätzer, this week called for closure of one-third of the clinic to stem mounting costs.

Sarah Toozee

Star wars

Sceptics abound one year on

Washington

PRESIDENT Reagan's request a year ago for the help of the scientific community in developing a system of defence against nuclear attack received a distinctly unfriendly answer last month. The Union of Concerned Scientists, (UCS), based in Cambridge, Massachusetts, marked the first anniversary of the President's "star wars" speech by publishing a detailed technical report saying the plan to create a comprehensive defence could not work.

Although UCS made plain a year ago its opposition to star wars, the new report is the most detailed critique yet published by opponents of the administration's strategic defence initiative. Compiled by a distinguished panel, including the physicist Hans Bethe, retired admiral Noel Gayler, defence analyst Richard Garwin and the physicist Victor Weisskopf, the report concludes that the prospects of a successful defence are minimal. Trying to develop one, however, would stimulate a new offensive round in the arms race, undermine the 1972 Anti-Ballistic Missile (ABM) Treaty and increase the risk of war.

The conclusions of the study reflect a growing consensus within the Department of Defense (DoD) itself that the president's ambitious objective in his so-called strategic defence initiative — to render nuclear weapons "impotent and obsolete" — is technically unattainable. DoD continues to support the initiative, but now says that although the new system would be useful to blunt a Soviet first-strike, it could not defend the civilian population.

According to the UCS study, the search for perfect protection is doomed from the start, because of the vulnerability of defensive facilities based in space and the absence of weapons that would be proof against Soviet countermeasures. The study predicts that the Soviet reaction to the development of a "star wars" system would include the development of a new generation of missiles, such as submarine-launched cruise missiles that could not be intercepted from space. Existing intercontinental ballistic missiles (ICBMs) would be fitted with more powerful engines, so their boosters would burn out quickly inside the atmosphere, where they would be less vulnerable to attack by X-ray lasers. Cheap decoy boosters without warheads could be deployed to overwhelm any space-based defence. And simple measures would be devised to knock out space-based defensive facilities in advance of a nuclear attack.

The study also dismisses as ineffective many of the technologies upon which a star wars system would have to rely. It says particle beam weapons are the least promising of the potential weapons, because charged particles could not penetrate the atmosphere. Chemical laser

weapons would have to be kept in low Earth orbit to be accurate enough, and — thanks to Newton's laws of motion — more than a thousand laser battlestations would be needed to keep a big enough fraction above Soviet silo fields all the time. Nuclear pumped X-ray lasers cannot penetrate the atmosphere, the study says. They could in any case deliver only a light blow from which ICBMs could be easily protected. Excimer lasers are dismissed as a

Yellow rain

Thai bees' faeces found

Boston

TWO US scientists have reported that there is now direct evidence that South-East Asian "yellow rain", which the US Government claims is a chemical warfare agent sprayed by the Soviet-supported regimes of Laos and Kampuchea, is actually a natural phenomenon — the defecation *en masse* by wild colonies of honeybees. Harvard biochemist and chemical warfare expert Matthew Meselson and entomologist Thomas Seeley of Yale University recently returned from an expedition to Thailand, during which they observed swarms of *Apis*, the true honeybees, making brief but abundant defecation sorties. The faeces come down as sticky yellowish spots up to a few millimetres in diameter which dry to a powder. They say that in composition and appearance they closely resemble samples of yellow rain from Laos that are the primary physical evidence for US allegations against the Soviet Union.

At the Khao Tai National Park in southern Thailand, the researchers say they found a defecation swathe extending as far as 160 metres from trees containing honeybee nests. Vegetation around the nests was covered with as many as 1,200 yellow spots of bee dung per square metre.

The significance of this discovery goes beyond biological observation. Since former Secretary of State Alexander Haig's Berlin speech of September 1981 in which he accused the Soviets of carrying out chemical warfare in South-East Asia, the US Government has maintained that samples of yellow rain collected in Laos provided by Laotian refugees are physical evidence of chemical attacks. These samples apparently contain tiny quantities of fungal toxins called tricothecenes produced by the *Fusarium* genus. This evidence has been a cornerstone of official allegations that the Soviet Union does not live up to the treaties it has signed. Use of chemical weapons would violate the 1925 Geneva Protocol which outlaws the use of chemical arms, and the 1972 Biological Weapons Convention. The US Government's case has been weakened, however,

"laboratory curiosity".

Last year's administration report by James Fletcher, former director of the National Aeronautics and Space Administration, maintained that even an imperfect defensive system would help the United States by reducing the confidence with which the Soviet Union could launch a disarming first strike. The UCS panel disagrees. It says a defensive system that posed a serious threat to the "assured destruction" capability of either side would result in a higher proportion of missiles being targeted on cities.

Peter David

by a growing body of evidence that supports the bee faeces theory.

Meselson believes he has an answer to the question of how tricothecene toxins got into the environmental samples of yellow rain and biochemical samples taken from ill Laotian refugees said to be victims of chemical attack. He notes that the reports of both yellow rain and chemical attacks have almost all occurred during an eight-week period between February and April, at the end of the dry season in tropical Asia. Infestations of *Fusarium* mould in food stocks are a particular problem at the end of the dry season in India, and Meselson says that mouldy sorghum is a problem in Thailand too. He explains the illness among the Laotian refugees and the presence of *Fusarium* toxins as resulting simply from naturally contaminated food.

Meselson and Seeley brought back samples of both faeces and local food supplies to test for mycotoxins. Since there has been considerable dispute over the quantification of mycotoxins in the yellow rain samples tested so far, Meselson intends to improve the methodology by adapting high resolution mass spectrometry to the analysis of tricothecenes.

It is still a mystery why the honeybees carry out the defecation flights. Honeybees in temperate climates are known to make "cleansing runs" on the first warm days of spring to purge themselves of faeces built up during hibernation. The synchronized behaviour of these tropical bees may reflect the presence in the hive of distinct foragers and defenders. When squeezed, defenders turn out to be full of faeces. Foragers are comparatively empty. **Christopher Earl Stephen Budiansky adds:** Last week, the State Department dismissed Meselson's latest findings as irrelevant to the issue of chemical weapons use in South-East Asia. Colonel James Leonard of the State Department said that the bee theory does not explain why mycotoxins were found on three samples that do not contain pollen, and that the US case rests not just on scientific evidence, but on looking at that evidence in "the fullest context" of intelligence data and refugee reports. □

National Science Foundation

Engineering lobby wins token

Washington

A LONG-STANDING campaign by Representative George Brown (Democrat, California) to force the National Science Foundation (NSF) to give greater attention to engineering was symbolically endorsed by the House of Representatives Science and Technology Committee late last month when it voted to amend the NSF charter to give explicit recognition to the agency's role in supporting engineering research. The move has revived an old debate over NSF's role and concern that science may suffer from any deviation from what critics such as Dr Frank Press, president of the National Academy of Sciences, call the "fundamental mission" for which the agency was created.

The committee's action, however, according to a committee staff member, in fact does little more than acknowledge the new facts of recent years — NSF's budget for engineering has grown from \$93 million in 1982 to \$147 million in President Reagan's 1985 budget request.

Brown's original proposal would have taken the symbolism one step further by changing the name of the foundation to the National Science and Engineering Foundation, but objections over the cost of reprinting stationery seemed to carry the day.

The Science and Technology Committee also recently decided to go one better than the administration by voting to triple the \$20 million that the administration had allotted to NSF for a new programme in the area of supercomputers. The committee was strongly influenced by testimony from researchers and by an internal working group study presented last summer to the National Science Board (the governing body of NSF) that called for hundreds of millions of dollars to be spent on establishing supercomputer centres around the country to support academic research. Research supercomputers are now confined to the weapons laboratories, some private industry and two special-purpose federal facilities, one for atmospheric research and one for magnetic fusion computing. The Reagan Administration plan stops short of purchasing new machines; the \$20 million would pay for computer time and access costs on existing machines. The House committee would leave it to an NSF advisory committee on supercomputing to decide just how the extra \$40 million would be spent, although it specifically encouraged acquisition of new machines, establishment of a national network and support for local access equipment and software development.

Stephen Budiansky

Unesco

Australia mounts defence

Canberra

AUSTRALIA is solid on Unesco (the United Nations Educational, Scientific and Cultural Organization), to judge from what the government and its ambassador to Unesco, former prime minister Mr E.G. (Gough) Whitlam have been saying. Indeed, the Australian National Commission for Unesco has urged the government to persuade the United States to reverse its stated intention of withdrawing from Unesco, deploring US intellectual isolation at a critical time in the search for world peace.

Elements of the Australian scientific community attach great value to many Unesco programmes, notably in the fields of hydrology, oceanography and ecology. Australia participates in several Unesco regional networks, which are seen as useful instruments of international cooperation, especially in Asia and the Pacific.

Mr Whitlam spelled out his own position here on 21 March, saying that while Australia supported the need for the press to be free, Unesco's work programme for 1984-85 on the New World Information and Communication Order (NWICO) had been accepted at last year's conference without dissent. Yet the programme is cited as one of the principal reasons for the pro-

posed US withdrawal. In any case, Mr Whitlam said, the privately-owned press is not as accurate as it should be.

The ambassador claimed that nothing in his own public utterances could be construed as an attack on the United States. In a partly-leaked version of a despatch to the Australian Government from his office in Paris, however, he is reported to have accused President Ronald Reagan of withdrawing from the organization as a sop to the "rabid right".

Mr Whitlam is further reported to have described Mr Amadou-Mahtar M'Bow, director-general of Unesco, as an outstandingly well-informed, competent and eloquent person: thoroughly alert, diligent and temperate. He recommended that the Australian Government invite Mr M'Bow to make an official visit to Australia, and defended the much-maligned Unesco staff, if defend is the right word, by allowing that they did not compare badly with Australian politicians, public servants and academics of his acquaintance.

One of Mr Whitlam's complaints was that the press had not mentioned that Unesco's books are normally certified by the British auditor-general. The US investigation of Unesco will begin this month.

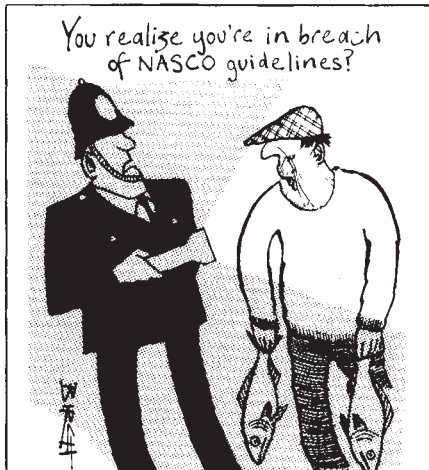
Jeffrey Sellar

Fishery stocks

Group leaps to aid salmon

THE North Atlantic Salmon Conservation Organization (NASCO) is now a going concern. But the inaugural meeting in Edinburgh earlier this year served simply to emphasize yet again that the migratory salmon poses the most difficult of all fisheries problems. The aim of NASCO is the conservation and also enhancement of salmon stocks in the North Atlantic. Its members are Canada, the United States, Denmark (in respect of the Faroes and including Greenland), Iceland, Norway and the European Economic Community (EEC) on behalf of interested member states — the United Kingdom, Ireland, Denmark and France.

The UN Conference on the Law of the Sea has laid down the principle that "states in whose rivers anadromous stocks originate shall have the primary interest in and responsibility for such stocks" and advocates a 200-mile limit, but the treaty is not yet in force. Meanwhile, EEC has prohibited salmon fishing in Community



waters beyond a 12-mile limit and sets quotas (2,000 tonnes in 1984) for Danish and West German fishing in the feeding areas off Greenland and in the Baltic. The Faroes have a separate quota agreement with EEC. Figures published by the International Council for the Exploration of the Sea (ICES) show a growth of Faroes/Danish catch in 1979-81 from zero to 1,025 tonnes.

NASCO will use ICES as a scientific advisory body and will try to formulate and rationalize international agreements to balance the needs of the international fishing community with those of individual countries.

Salmon frequently play a key role in the economies of rural areas poor in resources. Freshwater fishing is the largest participant field sport in the United Kingdom and France, for example. The Swiss are attempting to restock the Rhine with Swedish salmon. Countries of origin invest considerable sums in the conservation of

salmon and there is a need for international agreements to ensure that this commitment by the countries of origin sees an adequate return. Figures for Scottish rivers show a decrease of more than 50 per cent in the catches by rods and nets between the end of the 1970s and 1983, suggesting that the fish are not maintaining the breeding population.

It is not yet clear how far this decline is due to fishing in international waters or to over-fishing at home. Most British law on the subject is now more than 100 years old and takes no account of modern fishing

methods or the pressures to which the fish are subject. Meanwhile, the salmon farming industry has developed rapidly, particularly in Norway. In the United Kingdom, farmed salmon now account for 49 per cent of sales in a market that has grown 70 per cent in six years. This is rapidly changing the economic parameters, so that international fishing and illegal netting, a major law enforcement problem, will become less attractive. National measures will be meaningful only within the context of international control as envisaged under NASCO.

Sarah Tooze

Soviet science

Academic apologist emerges

OVERCROWDING and lack of equipment are seriously hampering the work of Soviet research institutes, *Pravda* claimed last week. The author of the article, although identified only as M. Korolev, clearly emerges as an apologist for the academic community in the long-running debate on why the results of Soviet research are implemented only slowly. Until recently, the standard explanation was that the scientists were failing to work sufficiently closely with industry. But lately there have been hints that the scientists are no longer prepared to shoulder all the blame.

During the late 1950s and early 1960s, there was what Korolev describes as an "explosive" expansion of research institutes in the Soviet Union which was fully justified by the demands of the national economy and the need to ensure that the Soviet Union could devise an independent solution to any technological problem but not enough attention was paid to the need that the institutes should be properly equipped.

Korolev also says that the various educational ministries had a very free hand, often founding institutes or upgrading them into universities simply for reasons of prestige. Constraints were imposed only later. Although these restrictions have reduced the great flood of new foundations, the Soviet Union still has a large number of ill-equipped and overcrowded institutes.

Korolev's anecdotal evidence suggests that prestige still plays a considerable role. Plans for a new building are passed, approved and included in the five-year plan but, before construction gets under way, the funds may be reallocated. Institutes are built, but equipment is not provided and, as at the Institute of Microbiology of the Byelorussian Academy of Sciences, scientists are left to "travel the whole country, from Grodno to Magadan", hunting for equipment on their own initiative.

Korolev claims many high-ranking scientists have no proper working space — even a laboratory chief at the Ufa petroleum institute complained of being without desk or chair although the situation is

apparently even worse in the humanities. (In the Moldavian Academy of Sciences, says Korolev, 70 per cent of specialists in the humanities have nowhere to work.)

Many of the worst cases cited by Korolev come from outlying republics. Institutes of the Turkmenian Academy and the Dagestan branch of the Soviet Academy, he says, have been waiting years for promised buildings, but even institutes in Moscow have their problems. The Institute of Earth Sciences of the Soviet Academy is scattered over 16 locations in the city, and working space is calculated as 2.7 m² per head. Development plans for Moscow are based on the concept of 342 scientific institutes in the capital, with a work force of 135,000. Is this, Korolev wonders, really the "optimum variant"?

Moreover, even if important institutes do manage to get the proper equipment, support facilities are not always available. Some 40 per cent of the scientific staff at the North Caucasus observatory, which boasts the largest optical telescope in the world (600 cm) still do not have their own apartments. (The foundations have been laid, but there is no money to pay the builders.)

Korolev's tentative solutions, given the apparent scale of the problem, seem somewhat unconvincing. Small institutes can be combined into "science centres" serving local industry (a scheme which was launched in 1975, but could be expanded further). Available resources should be used more rationally (he cites the case of computers used only three or four hours daily). A new drive for efficiency and production can be launched in the instrumentation industry. Research bases could be shared by more than one institute or by academic institutes and industry (a few such joint projects have already been launched).

Perhaps the most significant feature of Korolev's analysis, however, is the new awareness that the scientific workforce can no longer be expected to comply with high-level directives to "intensify scientific and technical progress", without the provision of at least a minimum of working facilities and basic living conditions.

Vera Rich

Polish universities

Elections cause some fears

UNIVERSITY autonomy and self-government implies that the universities must support "the goals of the socialist state", the Polish Government's press spokesman, Mr Jerzy Urban, said last week. The state, said Mr Urban, "cannot relinquish a firm influence over the character and the ideological content of university education", since it is university graduates who, in the future, will run "the country, the universities, everything. . ."

The issue of university autonomy was raised by Mr Urban in his weekly statement for the foreign press. According to Mr Urban, the elections of university and college academic councils now under way, and later on the performance of the new university authorities, will reveal whether the academic communities are willing to act in accordance with the letter and spirit of the Higher Education Act, and to abide by the "social accord" between Polish society and the government on which, he implied, the act is based.

The "social accord" has been a major theme of General Jaruzelski's speeches since martial law but, in spite of official claims, it seems not to enjoy nationwide support. Nor, with the act drafted during the Solidarity era still in people's minds, is there unanimity about what the letter and spirit of the Higher Education Act really are.

Mr Urban, it seems, does not expect the university elections to go through without problems — he said that "people who are firm and open opponents" of government policy "want to play a role in governing the universities" and the elections "may not be without elements of political struggle".

How the government intends to assert its authority if its political opponents are elected to university rectorships and councils did not need to be spelled out. The law introduced on 21 July 1983, when martial law formally came to an end, introduced amendments to the Higher Education Act, making it incumbent on university lecturers to teach "in accordance with the constitution of the Polish People's Republic", and providing for the dismissal of lecturers if their "political influence" on their students conflicts with this end.

Urban denied, however, that there had been any recent police actions at the Jagiellonian University of Krakow. The police had not searched university premises, he said, only the private homes of some faculty members suspected of underground activities. Moreover, reports in the Western media of road blocks around Krakow during the search had grossly misinterpreted the facts. The road blocks, he explained, were part of a separate action against black marketeers.

Vera Rich

British universities

Grants body answered

THE Royal Society has strongly criticized the way the UK Department of Education and Science (DES) is planning its expenditure on higher education. In a published response to a questionnaire on its views on the future of higher education, the society says it is "disturbed" that spending plans seem to be based on an uncorroborated student demand projection that is lower than is likely to be the case.

A wide-ranging series of questions on the future of higher education in Britain was circulated last November by Sir Peter Swinnerton-Dyer, the chairman of the University Grants Committee (UGC) after a request from the Secretary of State for Education and Science for a full debate on how higher education should develop. The questions were criticized in many quarters for appearing to assume that a substantial reduction in the level of expenditure per student is in prospect.

Of the several published estimates of demand for university places to the end of the century, the Royal Society's are the lowest. They show demand as roughly constant from now until the end of the decade, falling by 18 per cent to 1995-96 and rising thereafter. These estimates do not take account of the growing female participation rate, which is likely to increase demand for higher education. But the estimates of maximum total student demand promulgated by DES are below those thought plausible by the society, and DES appears now to be budgeting on its minimum demand projections.

Sir Peter Swinnerton-Dyer will not therefore be greatly surprised to learn that many of those replying have balked at being asked to consider what would be the consequences of a 1 or 2 per cent reduction in the level of financial support per student. The University of Technology at Loughborough probably speaks for many when it says in reply to this question that "there is a natural tendency to shrink from answering...on the assumption that any institution which thinks it could survive even the 1 per cent per annum reduction could well be asked to prove it".

Another contentious issue raised by the questionnaire concerns the earmarking of funds for research. At present, most of the universities' recurrent grant, paid by UGC, is for them to dispose of as they wish. But it has become an open secret in recent years that the cuts in real terms to UGC grants have compromised the universities' research capabilities, and the Medical Research Council has even gone so far as to make special provisions for university-based research groups that were being adversely affected by UGC cuts. It has therefore been suggested that UGC might specify what proportion of universities' grant is to go to research, even to the level of earmarking allocations to individual

research projects.

The Royal Society is ambivalent on this question. It proposes that the existing equipment grant, a component of recurrent grant already largely earmarked for science, should be increased to cover all the essentials of the "well-found laboratory". By thus expanding the scope of the equipment grant, UGC would have a firmer hold on the pattern of research as a whole, but would not be saddled with responsibility for making decisions between individual projects.

Deep-sea drilling

United Kingdom to miss out

THE United Kingdom seems unlikely to participate in the Ocean Drilling Project (ODP), the new phase of deep-sea drilling scheduled to start at the beginning of next year. The situation has arisen from an unfortunate combination of an increase in cost compared with the previous drilling programme, a lack of financial support from the Department of Energy and the inability of the Natural Environment Research Council (NERC) to make good the shortfall of about \$1 million.

Since the beginning of the Deep Sea Drilling Project in 1968, the research ship *Glomar Challenger* has travelled 96 "legs", drilling more than 1,000 deep-sea cores. These have yielded information not only about the structure and evolution of sediments and ocean crust but also about climate, ocean chemistry and circulation in the past. At the start of the International Phase of Ocean Drilling (IPOD) in 1975, the annual cost of participation to member countries was \$1 million, when 70 per cent of the UK contribution was provided by NERC and the Advisory Board for the Research Councils (ABRC) and the remainder by the Department of Energy.

In the new phase of ocean drilling now being planned, several countries outside the United States have expressed an interest by the payment of \$200,000, permitting involvement in the planning stage. The United Kingdom has paid this sum, as have Canada and a consortium including Italy, Norway, Sweden and the Netherlands. France and Japan are also involved and are intending to become full partners, while the Federal Republic of Germany has already done so. By being a full partner, a country will be able to participate in the planning and on-board operation of individual legs and will have immediate access to the samples stored in core libraries in the United States.

The cost of full membership of ODP will be \$2.5 million a year, an increase of \$0.5 million over that of IPOD in its final year. The increased charge results principally from the cost of operating a new ship: the

A similar compromise is suggested by the Science and Engineering Research Council, which suggests that universities be invited to submit research plans which UGC would take into account in fixing grants. The council declares itself willing to follow that road, presumably confident that any such tightening of controls on the amount spent on research could work only in its favour. The Advisory Board for the Research Councils has recently published warnings of dire consequences for British science unless the level of financial provision is subsequently increased, and has now decided to embark on an international comparative study to gather ammunition for its case.

Tim Beardsley

Sedco/BP 471 owned 50:50 by Sedco Inc. of Dallas and BP. The new ship has more sophisticated drilling capability, better weather tolerance and twice the scientific crew capacity of *Glomar Challenger*.

In 1983, the Department of Energy was supporting deep-sea drilling to the tune of more than \$0.5 million. Although no final decision has been made concerning ODP, the signs are that the department is very unwilling to provide funds for the project. According to a spokesman, the department received considerable benefit from IPOD, but without a firm programme a question mark remains about the relevance of ODP to the department's needs.

Scientists involved with ODP express a mixture of anger and despair at this position. They emphasize that the first 2½ years of the project will see the exploration of the Labrador Sea and the Norwegian Sea — both areas closely related in geological terms to the Rockall region where the Department of Energy's interests have been mainly directed. They suggest that the department is being shortsighted, ignoring the strategic value of ODP's programme to the oil companies. Caustic remarks are made about the hole drilled in the Rockall area under the management of the department at a cost of £25 million, which, according to one scientist, revealed "virtually nothing of value". Supporters of ODP also emphasize the planned drilling in the Weddell Sea — an area of considerable British interest.

The task now is to persuade the government and the oil industry to make more money available. One strong supporter of ODP is Sir Peter Kent, one-time chairman of NERC and ex-chief geologist of BP, who points out the disparity between the cost of the project and the huge taxes paid by the oil companies on North Sea oil recovered with the help of ODP's predecessor. The oil companies are not slow to notice this disparity either, and are considering the situation through the UK Offshore Operators Association.

Phillip Campbell

Solar physics

Shuttle repair a mixed blessing

Washington

SOLAR physicists are to be the chief beneficiaries of the first attempt by the National Aeronautics and Space Administration (NASA) to use the space shuttle for repairing a satellite in orbit. If this week's mission goes according to plan, astronauts will rendezvous with the crippled Solar Maximum satellite, which failed only months after its 1980 launch, and replace several defective components that are to blame for instrument malfunctions and the satellite's failure to keep an accurate lock on the Sun.

If the repairs — which will require the astronauts to spend two work days outside the shuttle — are successful, the satellite will once again provide solar physicists with data that are not available from ground-based observatories. The Solar Maximum satellite carries seven instruments, including ultraviolet, gamma-ray and hard X-ray detectors that measure radiation from the hottest regions of solar activity. These short wavelengths, which correspond to plasma temperatures of as much as 20 million kelvin, are cut off by gases in the Earth's atmosphere; ground-based observatories can "see" plasma temperatures only up to roughly 10,000 kelvin.

Although the satellite was designed principally to provide data on solar flares during the 1980 peak of solar activity, solar physicists say that there will still be much for Solar Max to look at. Solar flares are less frequent now; but the largest flares tend to occur during the declining part of the cycle, as now. The satellite will also provide data on active regions of the Sun in general and on other forms of activity, such as solar prominences and ejections of mass from the Sun. Successful repair of Solar Max's coronagraph — in effect, an optical telescope — may provide the added bonus of pictures of Halley's comet in 1986 as it passes near the Sun.

Only three of the instruments are at present providing data — a gamma-ray detector, a hard X-ray burst detector and an instrument that measures total solar output (which has demonstrated variability in what had been thought to be the constant solar constant). The failure of the satellite's pointing mechanism, which is supposed to keep it locked onto the Sun, has been the principal source of trouble; the satellite is now spinning up to 10° off axis in what NASA officials euphemistically term the "coarse pointing mode". The other instruments depend more critically on accurate alignment with the Sun; what is more, the coronagraph and the X-ray polychromator are suffering from additional malfunctions which the astronauts will attempt to correct.

The total cost of the repair mission, according to NASA's Goddard Space

Flight Center, will be \$45-50 million, which includes the cost of developing special tools and of training the astronauts. The Solar Max originally cost \$75 million; Goddard officials say its replacement cost today would be \$240 million. These figures do not however, include the cost of the shuttle launch.

Although the solar physics community generally supports the repair mission, some space scientists question the economics. Herbert Bridge, for the Center for Space Research at Massachusetts Institute of Technology, said last week, "I'm sure if they did a realistic cost estimate, the true figure would be a couple of hundred million. How you justify this I just don't know." Bridge did acknowledge the value of the repair mission as "a training exercise".

Other space scientists seem more genuinely enthusiastic. The solar division of the American Astronomical Society passed a resolution of support and Andrea Dupree of the Smithsonian Astrophysical Observatory said that both the immediate pay-off for solar physics and the development of repair techniques that can be used on other scientific satellites "are of equal weight".

The repairs, if successful, are expected to hold for between one and three years. As part of the operation, the satellite will be lifted from its current orbit of 270 nautical

miles to 290.

Future solar observations from space are now being planned but will be limited compared with Solar Max's capabilities. Spacelab 2, now scheduled for launch during 1985, will carry instruments to observe the "quiet Sun"; Spartan number two — a scientific payload that will co-orbit with a future shuttle flight — will measure the solar wind close to the Sun; and the large Solar Optical Telescope, a very high resolution, one-and-a-quarter metre telescope, is to be flown from the shuttle around 1990.

The Solar Max repair mission is a first for NASA, and will involve a number of untried procedures. The trickiest part may be the initial grab of the satellite, which is spinning at about one degree per second. After the shuttle approaches to within 200-300 feet, astronaut George Nelson will fly out using the jet backpack tested during the last shuttle flight. Nelson will approach the satellite in an arc, matching its spin, then grab onto it and halt the spin using his backpack thrusters. The shuttle will then approach and grab the satellite with its remote manipulator arm. Over the next two days, astronauts will repair the coronagraph's electronics module and the satellite pointing system and install a baffle on the X-ray polychromator, which has been plagued by solar plasma flowing into a vent port and saturating the instrument. If the operation fails, the entire satellite will be loaded into the shuttle cargo bay and returned to Earth.

Stephen Budiansky

Plagiarism accusations at Stanford

A FACULTY ethics committee at Stanford University Medical School is investigating charges that the chairman of the Department of Medicine, Dr Kenneth Melmon, plagiarized material from a medical textbook.

A chapter written by Dr Melmon for *Textbook on Endocrinology* contains 15 pages verbatim from *The Pharmacological Basis of Therapeutics*. Dr Melmon said he believed the editor of the endocrinology text had permission from the publishers of the pharmacology text to use the material.

Although both books were published in 1979, it was only last month that Macmillan Publishing Company of New York complained to Stanford that its pharmacology book had been plagiarized.

Dr Melmon is "bewildered, in shock and dismayed" by the accusations, said Stanford law professor Jack Friedenthal, who has volunteered legal services to the physician. "This is one of those agonizing, unfortunate things that could happen to any one of us as we all use one another's work in legitimate ways."

Dr Melmon's problem is that the only witness to his side of the story, Robert H. Williams, is dead. According to Dr Melmon, he included the section on

pharmacology at Dr Williams's urging. Dr Williams, the editor of the textbook, said he would arrange to get the necessary permission, and later telephoned to say that permission had been granted.

Dr Williams has since died of a heart attack, and Dr Melmon has no proof to support his contention that permission was sought and granted.

None of the editors who worked on the pharmacology text recall or have records of anyone asking permission to use the material in question.

According to Mr Friedenthal, Dr Melmon would have nothing to gain by plagiarizing the material. Sceptics say that people who plagiarize seldom think they will get caught.

A medical school standing committee on ethical scientific performance is looking into the matter and will make a report to Stanford president Donald Kennedy in about a week. It must decide if Dr Melmon is telling the truth and what is to be done about the charges from the publisher.

"This has been very painful to Melmon", Mr Friedenthal said. "There is a cloud over his head and suddenly his fine reputation has been dragged through the mud."

Sandra Blakeslee

Nuclear disarmament

SIR — To be successful, any scheme for multilateral disarmament should proceed by small steps, should convince both sides that each has bettered the other and, most importantly, should not become embroiled in the difficulties of weapon comparison. I propose a mechanism which has these three characteristics. Indeed, it can turn to advantage the inevitable differences of opinion about weapons of the two sides. It is based on the "I cut — you choose" rule by which children can divide a cake.

Each side begins by assigning a number to each separable nuclear device in its armoury. This number, the "military value percentage", is chosen by the weapon owner to represent his view of the usefulness of the item as a part of his entire inventory. The sum of all the numbers of each side is equal to one hundred. To take an example, if the Soviet Union decided that the 350 missiles in the SS-20 system represented, say, 15 per cent of its nuclear strength then the military value of each would be 0.04292 per cent.

It would be extraordinary if the values of usefulness chosen by one side were in exact agreement with the magnitude of threat felt by the other. Indeed, we may expect that the weapons with accurate terminal guidance and short launch times, which are suitable for pre-emptive first strikes, will induce a feeling of threat in their victims which is much greater than the feeling of comfort they offer to their owners. On the other hand, second-strike weapons are valuable deterrents and provide a large feeling of security, but do not pose a threat in proportion. This difference of opinion provides the incentive for the disarmament process and ensures that both sides can believe that they have secured advantage.

The first reduction should be very small. Let us suppose that it is a step of about 1 per cent. Each side picks from the list of its opponent the most threatening items with total military value percentage not exceeding this "table limit". The selections may be announced simultaneously and small differences carried forward as credits for a second round.

If the United States happened to decide that the SS-20 was the most serious threat, they would request as a first move that the number of missiles be reduced by 23. Meanwhile the Soviets would pick the most threatening 1 per cent of weapons from the US list. The United States would be quite indifferent about the Soviet choice because the numbers would have been chosen to make any 1 per cent selection equal, in their view, to any other.

Both sides will think they benefit from this exchange by an amount which depends on the ratio of perceived threat removed to perceived protection lost. The absolute, as opposed to the relative, magnitude of the reductions of each side, measured in terms of fire-power or lethality, will be greater

for the power with the greater original armoury. But as both sides argue that the other has the excess they can hardly object to this feature of the scheme.

The problem of verification is common to all disarmament plans. A necessary assumption for any scheme is that both sides have reasonably accurate knowledge of the weapon systems of their opponents. If the reductions proceed by small, slow steps, then neither side need fear that its national security has been greatly endangered if verification goes wrong. But if a side is sincere about its wish to disarm, it can use the interpretation of verification procedures to send messages about its sincerity and entice the other side to continue.

Either side may wish to distort the percentage values it declares. But because the sum total is always equal to one hundred a reduction is quite legitimate but the ploy may backfire and lead to the loss of good weapons at less than their true value. It is also possible to design rules which allow updating of weapons. For example, if Side A insists on the introduction of some new missiles, it may do so provided that it also declares a military value percentage for it. Side B may then, without loss to its armoury, remove items to that same value from any part of Side A's inventory including the new ones. Side A will not want the new ones to be instantly lost and so will have to put a higher than true value on them. It will therefore have to give up rather more of its obsolete inventory. This rule would encourage the evolution of new weapons which provide high perceived security for low perceived threat — a most desirable feature.

STEPHEN SALTER

*Department of Mechanical Engineering,
University of Edinburgh,
Edinburgh, UK*

Page charge revolt?

SIR — The problem of discrimination by journals against impecunious authors is a serious one. It is a moral issue as much as a practical one: should ability of the author to pay be as important as scientific merit in accepting a paper for publication? Most journals require both, so each has an equivalent veto.

A journal can argue that it can publish less if it has less income, can charge less to subscribers if it collects from authors, and that authors who can afford to pay page charges often do not do so when the charges are voluntary. These problems are real ones, but do they really outweigh the denial of the possibility of publication to those unable to pay for it?

It is not adequate to say, as a few journals do, that "in exceptional circumstances" page charges will be waived. Be-

ing impecunious is no longer exceptional, if it ever was. Neither is it adequate to require membership in a society even for authors marginal to it or unable to pay. And giving special treatment to those who pay of course again raises money to the rank of merit.

Publishers are not likely to give credit on page charges to referees (*Nature* 2 February, p.408), and anyway this would not help other impecunious authors. For the past decade I have neither submitted papers to, nor refereed for, any journal which practises financial discrimination.

It is often convenient to publish in discriminatory journals, and often these have good reputations scientifically. But if even those of us who can pay do not help such journals, and tell them why, we may be able to correct a serious iniquity.

LEIGH M. VAN VALEN

*Biology Department (Whitman),
University of Chicago,
Chicago, Illinois 60637, USA*

Opposite to placebo

SIR — L.J. Barrie (*Nature* 16 February, p.590) bemoans the lack of a term the opposite of a placebo: a harmless object producing a harmful effect if the recipient believes it to be noxious. Surely, the word is "antiplacebo". Some examples: the quite serious effect the bite of a harmless snake can produce if the victim believes it to be poisonous, or the bite of a poisonous snake when no venom has been injected into the wound. "Pointing the finger", curses, prophecies of evil, are of this nature. As a rule, the antiplacebo is not an object, like a pill; its potency is based entirely on its psychological effect, but so, in reality, is that of the placebo.

STEPHEN SEELY

*University of Manchester Medical School,
Manchester M13 9PT, UK*

SIR — "Garblow" is an unfortunate proposal for a word to express the opposite of a placebo. "Antiplacebo", used by the late Gregory Bateson to describe a cancer diagnosis, has the advantage of suggesting its meaning. Shorter and equally Latin would be "displacebo".

PHILIP J. STEWART

*Commonwealth Forestry Institute,
South Parks Road,
Oxford OX1 3RB, UK*

SIR — "Garblow" is an unnecessary dysphemism for "nocebo" to define "a harmless substance which produces harmful effects when the patient believes it to be toxic". "Nocebo" was introduced at least 15 years ago (G. Herzhaft, "L'Effet nocebo" *Encéphale* 58, 486; 1969).

M.J. Barrie also states that little attention has been paid to the harmful effects of placebos. In fact, Wolf (*Pharmac. Rev.* 11, 689-701; 1959) and Honigfelt (*Dis. Nervous Syst.* 25, 145-156; 1964) each cite 10 references to them. ARTHUR CHERKIN
*Veterans Administration Medical Center,
Sepulveda, California 91343, USA*

Alternatives to the Big Bang

Could the Big Bang be an illusion caused by a generation of massive stars formed before galaxies came into being? A new study suggests this is unlikely but not impossible.

AFTER a long period during which it has been customary to suppose that most stars in the Galaxy and elsewhere are formed with a mass not very different from that of the Sun, give or take a small-integer factor, very massive stars are again now all the rage.

Both fashions are understandable. Stars which are formed as objects much less massive than the Sun, objects such as Jupiter for example, will never yield a combination of temperature and pressure great enough to ignite thermonuclear reactions and so will be inconspicuous unless they happen to be part of the Solar System. But because the rate of thermonuclear reactions is exceedingly sensitive to the temperature at the core of a star, objects much larger than the Sun will run through their evolution more quickly, their larger bulk of fuel notwithstanding. So whatever the distribution of stellar masses in the Galaxy at its formation, the passage of time, a few Solar System lifetimes perhaps, will have ensured that the stars in the sky are roughly comparable in mass with the Sun. The anthropic principle is evidently applicable.

Although there is no compelling observational reason for paying more attention now than a decade ago to the properties of stars that are much more massive than the Sun, there are other reasons for learning a little more about their properties. So much is, for example clear from the argument put up by Professor M. J. Rees some years ago (*Nature* 275, 35; 1978) that there could have been a time in the early history of the Universe, before galaxy formation set in, when stars with masses 100 times or more the mass of the Sun could have had a necessarily brief existence. The point of that argument was to demonstrate that a population of stars like this, perhaps 1 million years or thereabouts after the Big Bang, would have produced radiation in an amount and of a character that would now simulate the 2.7 K microwave background radiation. The necessity of the Big Bang was itself weakened, but not undermined.

The cosmological implications of the possibility that very massive stars may be capable of existence has now been taken a long step further by B. J. Carr (a colleague of Rees at the University of Cambridge), J. R. Bond (Stanford University) and W. D. Arnett (University of Chicago). There are several ways in which stars like this may have influenced the course of events, of which the most appealing is that their rem-

nants, probably black holes, could provide enough unseen mass to hold the Universe together. But naturally, such stars could have contributed to nucleogenesis in the early Universe and perhaps even to primordial helium abundance (thus further weakening the case for the Big Bang). Writing in the *Astrophysical Journal* (277, 445-69; 1984), Carr *et al.* set out to tell what can be learned about the properties of early massive stars from the several different roles that have been suggested for them.

Much of the argument turns on calculations by the same authors (and not yet published) of the final state of massive stars, more than 100 times the mass of the Sun, for example, at the end of their evolution. It seems agreed that stars this massive will from the start be unstable, liable to disruption by fluctuations of the rate of thermonuclear reaction at the centre, even while they are burning hydrogen as fuel. But if they avoid catastrophe until the hydrogen in the central region has been converted into oxygen, further nucleosynthesis will be less probable than other more exotic reactions, and in particular the creation of electron-pairs (first predicted in 1964 by W. H. Fowler and F. Hoyle). Carr *et al.* now say that what will happen physically to the star will depend on the mass of the unburned core of oxygen; if this should exceed 100 or be less than 30 solar masses, the object will collapse into a black hole but will otherwise explode.

With this information, it is possible to ask sensible questions about the early population of massive stars. If, for example, their original mass distribution was the kind of distribution expected of matter free adiabatically to agglomerate, and if their remnants (black holes) now provide the missing mass that holds the Universe together, something can be said about numbers and mass distribution. But then the formation of the whole array of black holes at this early (and uncertain) epoch in the Universe would have triggered off a series of gravitational wave trains which, after the passage of the present lifetime of the Universe, would have a frequency spectrum calculable from the initial mass distribution of the massive stars. So, in principle at least, it should be possible to set limits on the numbers of these early collapsing objects from measurements now of the density of gravitational waves.

Exactly the same arguments apply to the

other phenomena on which early massive stars would be expected to have a bearing. The observed upper limits on visible background radiation in the Universe for example require that there should not have been too many massive stars, or that the epoch at which they existed should not have been too recent. Similarly, the microwave background radiation, to which radiation from early massive stars would have contributed directly (after allowing for the frequency reduction caused by the expansion of the Universe as well as scattering by grains between stars) provides another constraint which can in principle be tested by observation.

The most stringent constraints, as it turns out, are provided by the need that early stars should not so have populated the Universe with heavy metals that it is impossible to account for stars still shining which are virtually free from heavy elements. At least some stars are known which appear to consist almost exclusively of hydrogen and helium, and must therefore be supposed to have been formed from material virtually free from heavier elements. Since it seems to be agreed that stars with masses say 10 times that of the Sun would be, during all but the earliest stages of their evolution, prolific sources of elements such as magnesium (witness the calculations of K. Nomoto in the same issue of the *Astrophysical Journal* 277, 791; 1984), it seems to be necessary that most of the matter locked up in massive stars should have been in stars so massive that they would have collapsed once carbon and oxygen came to predominate in their core.

On balance, Carr *et al.* appear a little disappointed with the outcome of their work. Plainly it will be very difficult to account for cosmological phenomena as different as the known microwave background and the primordial abundance of helium and supposed near-absence of heavy elements by substituting a generation of massive stars for the now-fashionable Big Bang unless the character of that lost generation of stars is tightly, even artificially, constrained. But that will not inhibit these and other explorations of the notion that the Big Bang may not be the only plausible early state of the Universe. For, as Carr *et al.* have shown, the problem of massive stars is neither abstract nor intractable. And Carr *et al.* are asking that very massive stars should account for almost everything.

John Maddox

Archaeology

Absolute dating of the Bronze Age eruption of Thera (Santorini)

from Peter Warren

THE establishment of any absolute or calendrical date for a major event in Aegean prehistory would be a matter of great interest, not least because the Minoan and Mycenaean civilizations of that period exhibit many peaks of technical, social and artistic achievement. The recent, albeit tentative, dating¹ of the Bronze Age eruption of Thera to 1628–1626 BC is therefore certain to stimulate archaeologists just as the date² of 1390±50 BC stimulated them when it was put forward in 1980.

The newly suggested date is that recorded as frost damage in the bristlecone pine tree-ring sequence from Campito Mountain, eastern California; the link with the Thera eruption hinges upon the suggestion that, like some other major volcanic eruptions, it induced sufficient climatic change to leave a record of frost damage. The date of 1390±50 BC was based on a peak of acidity in the Camp Century ice core from north-west Greenland and is based on the assumption that acidity is a result of volcanic eruption. For either date to be correct it needs to be compatible with the archaeological dating of the eruption, which I shall summarize.

The Bronze Age settlement at Akrotiri on Thera, excavated from 1967 to 1974 by Sp. Marinatos³ and since by C. Doumas^{4,5}, came to an end at a time when it was full of Late Minoan I A and locally equivalent pottery. It is assumed by many that it was the eruption that destroyed the settlement. Others, including the writer, hold the view that evidence of reoccupation and clearance of parts of the destroyed settlement before the pumice fall suggests an interval of time between the end of the settlement and the eruption. In that case, the eruption could have occurred in the subsequent archaeological stage, Late Minoan I B, and be correlated with the greatest destruction of the major Minoan sites of Crete, which are characterized by Late Minoan I B pottery⁶.

Absolute dating of these ceramic periods from artefactual evidence is far less precise than we would like. Associations of Late

Minoan I B (and its Greek mainland equivalent, Late Helladic II A) with Egyptian material sets the period at the time of Tuthmosis III (refs 7,8), the fifth pharaoh of the 18th Dynasty, the highest dates for whom are 1504–1450 BC, the lowest 1479–1425 BC (ref.9). A probable correlation involving Crete, Taanach, near Megiddo in Israel, and Egypt indicates (on an accession date of 1504 BC for Tuthmosis III) a date of around 1482 BC for a point within Late Minoan I B^{7,8}. May the Late Minoan I B styles have existed before the time of Tuthmosis III?

In relation to this, a pottery fragment from Abydos tomb 328 in Egypt has recently been dated to Late Minoan I B. The surviving decoration on the fragment is 'adder mark' and dot motif and star motif (Fig.1), well paralleled in Late Minoan I B, yet most of the material in the tomb is datable to the Second Intermediate or Hyksos (pre-18th Dynasty) period (although an Egyptian pottery bowl found in the tomb is also paralleled in the 18th Dynasty, at the time of Tuthmosis III). Thus, while the weight of the evidence seems to support a pre-18th Dynasty date for the fragment of Late Minoan I B pottery, a date in the time of Tuthmosis III cannot be excluded. It is also the case that Late Minoan I B has never been thought to be a long ceramic period: there is almost no evidence for a stratified sequence of Late Minoan I B deposits and the period may well have been the product of just one or two generations of vase painters. Given the other Egyptian associations of Late Minoan I B it is not easy to see how the Abydos fragment could have had a pre-18th Dynasty context.

Late Minoan I A is conventionally dated 1550–1500 BC, beginning after the inception of the 18th Dynasty¹¹ and followed by Late Minoan I B. A date for the beginning of Late Minoan I A pottery has not yet been firmly established, and it may be earlier than 1550 BC. A number of gypsum vessels from the Akrotiri settlement on Thera (Late Minoan I A) may well

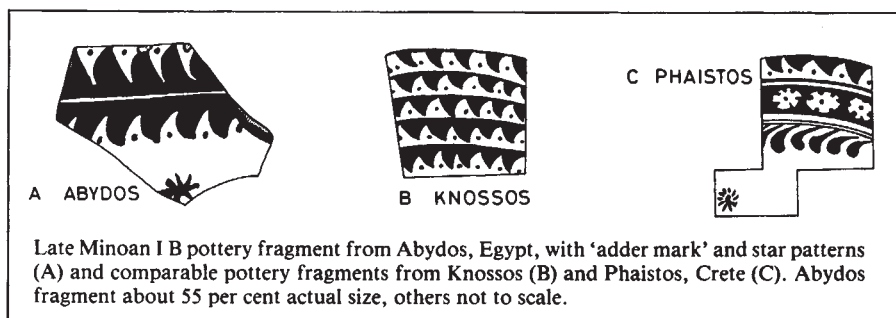
be Syro-Palestinian Middle Bronze II products, which implies a pre-18th Dynasty Second Intermediate period date in Egyptian terms¹². This would put Thera Late Minoan I A before 1576 or 1552 BC (alternative starting dates for the 18th Dynasty⁹). An upper limit can be inferred from the famous alabaster lid from Knossos, inscribed with the cartouche of the Hyksos pharaoh Khyan, which has been dated to Middle Minoan III (ref.13, though see refs 14,15) which preceded Late Minoan I A. Khyan ruled in Egypt within about 1660–1580 BC. So Late Minoan I A could well have begun by 1600 BC though the period postdates the apparent Khyan–Middle Minoan III correlation.

If, at first sight, the date of 1600 BC seems compatible with date of 1626 BC from the frost damage record¹, it must be remembered that 1600 BC is at the beginning of Late Minoan I A, while the Thera eruption is certainly after the beginning of Late Minoan I A (ceramics from that period are present in the buildings predating the main destruction buildings at Akrotiri) — in fact there is no known stage of pure Late Minoan I A later than the Thera destruction deposits, themselves linked by ceramics to at least six other Late Minoan I A deposits in Crete. Thus from the archaeological evidence the Thera eruption does not appear to be datable anywhere near 1626 BC or even 1600 BC. At present the balance of the archaeological evidence dates the end of Late Minoan I A no earlier than about 1500 BC. If the eruption is linked to the great Cretan destructions of Late Minoan I B, its archaeological date would be about 1450 BC, which is near the date of 1390 ± 50 BC established by Hammer *et al.*².

Radiocarbon dates from destruction contexts at Akrotiri^{8,10,16,17} also do not offer strong support for the 1626 BC date. Of the ten short-lived samples, six, after calibration, fall in the 17th century BC, one is a little earlier, another is 1490 ± 50 BC and two are well back in the third millennium. Moreover, they may have yielded dates older than a true date¹⁶.

Just how secure is the suggested correlation of the frost event at 1626 BC with the Thera eruption? First, according to Tables 1 and 3 in the paper by LaMarche and Hirschboeck¹, for 9 of the 19 defined volcanic events since AD 1500 there are no corresponding frost-ring events — that is, there is only about one chance in two that an eruption will induce notable frost-ring damage. Furthermore, of 17 listed frost events since AD 1500, 7 do not correspond to a listed major volcanic event.

Second, if we compare the dendro-chronological sequence (from Campito Mountain, White Mountains; Table 3), which produced the 1626 BC date, with other tree-ring data from western USA (Table 1), of the 26 listed notable frost-ring events (including the one at AD 1500) since AD 595, the White Mountains sequence appears to record only 7, while it produces



6 others not present at the other locations (perhaps because smaller sample criteria were allowed for Table 3, White Mountains). Of the 13 White Mountains frost-ring events, only 5 correspond to dated volcanic eruptions. Finally, of the 11 observable joint occurrences (frost-rings and eruptions) since AD 1500 (and counting the AD 1500 events as a joint occurrence) only 3 appear in the White Mountains frost-ring sequence.

Third, with 19 major volcanic events listed since AD 1500, extrapolating back over the complete White Mountains sequence (to 3435 BC), one would expect 201 volcanic events. Even if every listed frost-ring from the White Mountains were postulated to match a notable volcanic event (and clearly more than half do not) the record of just 14 pre-AD 1500 frost-rings from that source means that it would be picking up only about 1 in 14 of the postulated 201 volcanic events.

Fourth, frost-ring events are proxy data — they could be caused by volcanic eruptions; but they could equally well be caused by other influences on climate. By contrast, high acidity peaks seem to be produced specifically by volcanic eruptions. For the period 3435 BC to AD 1963/5, of the 23 eruptions producing high acidity peaks listed by Hammer, Clausen and Dansgaard, only about 10 correlate with frost events, while no less than 26 out of 36 frost events have no matching acidity peak. The frost event of 1628–6 BC has no matching acidity peak, the nearest being in 2690 ± 80 and 1390 ± 50 BC.

On this evidence, it would seem there is little support for linking the White Mountains frost-ring event at 1626 BC specifically with the eruption of Thera. Moreover the link is not substantiated by archaeological data, nor strongly supported by radiocarbon evidence from Thera. □

1. LaMarche, V.C. Jr & Hirschboeck, K.K. *Nature* **307**, 121 (1984).
2. Hammer, C.U., Clausen, H.B. & Dansgaard, W. *Nature* **288**, 230 (1980).
3. Marinatos, Sp. *Excavations at Thera*, 1 (Archaeological Society of Athens, 1968–1976).
4. Dumas, C. *Thera. Pompeii of the Aegean. Excavations at Akrotiri 1967–79* (Thames & Hudson, London, 1983).
5. Warren, P. *Nature* **304**, 469 (1983).
6. Page, D.L. *The Santorini Volcano and the Desolation of Minoan Crete* (Society for the Promotion of Hellenic Studies, London, 1970).
7. Hankey, V. & Warren, P. *Bull. Inst. class. Stud. Univ. Lond.* **21**, 142 (1974).
8. Betancourt, P.P. & Weinstein, G.A. *Am. J. Archaeol.* **80**, 329 (1976).
9. Kitchen, K.A. *Serapis* **4**, 65 (1977–78).
10. Kemp, B.J. & Merrillees, R.S. *Minoan Pottery in Second Millennium Egypt* (Zabern, Mainz-am-Rhein, 1980).
11. Evans, A.J. *The Palace of Minos at Knossos Vol.2* (Macmillan, London, 1928).
12. Warren, P. *Archaiologiki Ephemeris*, 82 (1979).
13. Evans, A.J. *The Palace of Minos at Knossos Vol.1* (Macmillan, London, 1921).
14. Palmer, L.R. *Nestor* **2**, 323 (1964).
15. Hood, M.S.F. *Nestor* **2**, 342 (1964).
16. Michael, H.N. in *Thera and the Aegean World Vol.1* (eds Dumas, C. & Puchelt, H.H.) 791 (Aris & Phillips, London, 1978).
17. Michael, H.N. in *Thera and the Aegean World Vol.2* (ed. Dumas, C.) 245 (Aris & Phillips, London, 1980).

Peter Warren is Professor of Ancient History and Classical Archaeology in the Department of Classics and Archaeology, University of Bristol, Bristol BS8 1RJ.

Materials science

A soft superionically conducting lower mantle?

from Robert W. Cahn

THE meagre contribution so far of materials science to the general understanding of the Earth's mantle is about to be enriched, in part by the study of ionic crystals whose properties may be a guide to those of the mantle rocks, with their capacity to undergo plastic deformation and with their large electrical conductivity.

Two lines of geophysical arguments have helped to define what needs to be explained. First, studies of post-glacial rebound, the gradual rising of the Earth's surface after the removal of overlying ice sheets, is a pointer to the viscosity of the convective mantle. Some, however, propose a uniform viscosity for the whole mantle, others, layered viscosity variations. Weertman¹ takes it as probable that the lower mantle is orders of magnitude more viscous than the mean mantle viscosity.

Second, the conductivity of the mantle, especially at its lower levels, is linked with the delays between changes in the magnetic field of the core and its manifestation at the Earth's surface — what magneticians call magnetic viscosity. Thus Anderson² has recently pointed out that the delay associated with a magnetic 'jerk' such as that of 1969 must be intimately linked with mantle conductivity, but has also drawn attention to the difficulties in putting a value on the mean conductivity from such observations. Nevertheless, Stacey³ has estimated, from such data together with an analysis of the secular variation of the length of the day, an electrical conductivity at the bottom of the mantle (itself orders of magnitude higher than in the upper mantle) of $100\text{--}500 \Omega^{-1} \text{ m}^{-1}$.

It is widely held that the major constituent of the lower mantle is a pressure-induced polymorph of pyroxene, $(\text{Mg,Fe})\text{SiO}_3$, with perovskite structure. The implication is that the mantle is chemically but not structurally homogeneous (though Whaler⁴ has recently advanced seismic arguments for presuming chemical heterogeneity).

In assessing the compatibility of the conventional model of the lower mantle with the values of viscosity and conductivity needed to fit data on glacial rebound and magnetic delay, the difficulty arises that $(\text{Mg,Fe})\text{SiO}_3$ of perovskite structure can be made in the laboratory only at such high pressures, and therefore minute volumes, that direct measurements of its properties are not feasible.

In any case, in spite of the large numbers of perovskite-type compounds known⁵, there has been no investigation whatever of the plastic deformation of any of them.

Indeed, Poirier *et al.*⁶ point out that there is no direct knowledge of the phase (*pace* Whaler) presumed to occupy 40 per cent of the Earth's volume, and have accordingly undertaken a study of KZnF_3 , a salt of perovskite structure which is stable at ambient pressure, up to its melting point of 870°C . In this they followed the lead of O'Keefe and Bovin⁷, who examined NaMgF_3 , another perovskitic salt, and found that it becomes a solid electrolyte (alias a superionic conductor) at high temperatures.

The special importance of the work of Poirier and his colleagues is that they examined the mechanical as well as electrical behaviour of KZnF_3 , showing both that it becomes a superionic conductor and that it acquires a low viscosity above the same critical temperature range. This remarkable finding is of importance both for geophysicists and for materials scientists.

Specifically, Poirier *et al.* have measured the creep curves of oriented crystal slices of KZnF_3 between 650° and 844°C and electrical conductivity between 650° and 800°C . Both primary and steady-state creep was observed. The latter has approximately newtonian characteristics, with a rate proportional to stress at a fixed temperature. At a fixed stress, the creep rate increases with temperature up to 750°C , decreases slightly between there and 800°C and then increases steeply at higher temperatures.

The conductivity rises slowly with temperature up to a value of $0.8 \pm 0.2 \times 10^{-2} \Omega^{-1} \text{ m}^{-1}$ at 700°C and then rapidly to $1.9 \pm 1.8 \Omega^{-1} \text{ m}^{-1}$ at 800°C , with most of the increase between 700° and 750°C . The values agree well with those reported for NaMgF_3 by O'Keefe and Bovin — but note the high degree of experimental scatter near 800°C .

Weertman¹ had already concluded that newtonian creep is necessary to account for observations of post-glacial rebound. Newtonian creep usually involves diffusion (Nabarro) creep in fine-grained materials, not in single crystals as in Poirier's work. Poirier *et al.* conclude that they have observed an example of Harper-Dorn creep, a mechanism reported many years ago but badly understood. They argue that the sharp increase of diffusivity of one ionic species, here F^- , associated with the advent of superionic conduction (see for example, ref. 8) should hinder creep by promoting the climb and immobilization of some dislocation segments in crystals of KZnF_3 . Consequently, dislocations should not multiply in the familiar way during

plastic deformation, but pre-existing dislocations should drift under stress according to the Orowan law, which specifies a drift rate proportional to stress.

Poirier *et al.* also discuss in some detail possible mechanisms for the hardening of the crystals between 700° and 750°C. They further point out that, once the crystals become superionically conducting, the F-sublattice effectively 'melts'⁹ and this entails a change in the core structure of dislocations to a B2, β -brass type, with resultant easy slip on {100} as observed.

From a geophysical standpoint, if perovskitic (Mg,Fe)SiO₃ behaves similarly to KZnF₃, and if it is indeed the major constituent of the lower mantle, then the viscosity of the mantle would not increase with depth, as has hitherto been assumed, but might actually decrease below ~700 km. The maximum conductivity of KZnF₃, about 2 $\Omega^{-1} \text{ m}^{-1}$ is still about two orders of magnitude less than the value deduced as consistent with superionic conductivity in perovskitic (Mg,Fe)SiO₃. But as the maximum conductivities of known ionic conductors range up to 500 $\Omega^{-1} \text{ m}^{-1}$ this disparity poses no problem.

To materials scientists, the interesting feature of these measurements is the occurrence, in a narrow temperature range, of steep changes in both viscosity and conductivity. A check through two major conference proceedings concerned with superionic conductors^{8,10} reveals no investigations whatsoever of the mechanical characteristics of any member of this family of materials.

There is, however, another family of crystals which shares with superionic conductors the characteristic that one sublattice only undergoes progressive disordering as the temperature is increased. These are the plastic crystals, organic compounds in which some of the molecules acquire both positional and orientational disorder and which are so called because, in the partially disordered state, they become extremely soft so that some will flow even under gravity. When that happens, self-diffusivity also becomes extremely high and it is widely assumed that the two changes are linked, but the nature of the relationship is not clear.

The general characteristics of positional and orientational disorder in many types of crystal, including ionic superconductors, and of the flow and self-diffusion of plastic crystals are all very clearly set out in an

excellent monograph by Parsonage and Staveley¹¹. While it is not likely that plastic crystals (molecular) and superionic conductors (ionic) have similar mechanical characteristics, the work of Poirier and his colleagues has nevertheless opened an entirely new field of research into the relation between the disordering or 'melting' of one sublattice in a crystal and

the consequential changes in diffusivity of either ions or molecules, and the further connection between these two features of the crystal and its mechanical softening. □

Robert W. Cahn was formerly Professor of Materials Science at Sussex University and Professor of Metallurgy at the University of Paris-Sud, and can be contacted at 6 Storey's Way, Cambridge CB3 0DT.

Superconductivity

Nitride offers 30K transition?

from D.A. Papaconstantopoulos, W.E. Pickett, B.M. Klein and L.L. Boyer

MATERIALS capable of making a transition to the superconducting state at a temperature significantly greater than the liquification temperature of hydrogen would have important applications in technology, for which reason it is disappointing that there has been no increase of the largest-known transition temperature, T_c , in the past ten years. Suggestions, on theoretical grounds, that mechanisms of superconductivity other than the interaction between phonons (lattice vibrations) and electrons could yield higher values of T_c have not so far materialized. In what follows, we argue that the B1 crystal structure of molybdenum nitride, MoN, is a prime candidate for a 'high-temperature' superconductor and that it should have a transition temperature in the region of 30K.

The quality of the numerical results of theoretical studies of the electron-phonon interaction and T_c in the past decade largely parallels the accuracy of energy-band calculations for metals and metal compounds (for reviews, see refs 1,2). The experience of our group and others suggests that the B1-structure carbides and nitrides are good candidates for materials with high T_c , and for two reasons. First, in these materials, the density of electronic states at the Fermi level, $N(E_F)$, is relatively high. Second, the electron-phonon matrix

elements are large, a reflection of the strength of metal-carbon and metal-nitrogen bonds. Both these attributes favour high T_c but they also lead to structural instabilities which make it difficult to stabilize the stoichiometric structures, or magnetic instabilities which quench superconductivity.

In the past few years, we have nevertheless calculated the band structure and the superconducting properties of a number of transition-metal carbides and nitrides with NaCl structure¹⁻⁵. Since these calculations are in good agreement with experiment, we are confident of our predictions by similar calculations for other materials in the same class which have not so far been produced in the laboratory. Recently we proposed⁶ that MoN in the B1 (NaCl) structure should be a superconductor with $T_c \sim 30\text{K}$.

The equilibrium MoN phase diagram⁷ reveals a hexagonal compound, δ -MoN, at a strict one-to-one atomic ratio which is stable below 800°C, but with available methods of preparation (sputtering, electron beam evaporation, shock compression vapour deposition and ion implantation) it is not unusual to form metastable phases such as B1-structure MoN. Given the predicted high value of T_c for this material, we believe that a concerted effort to induce its formation

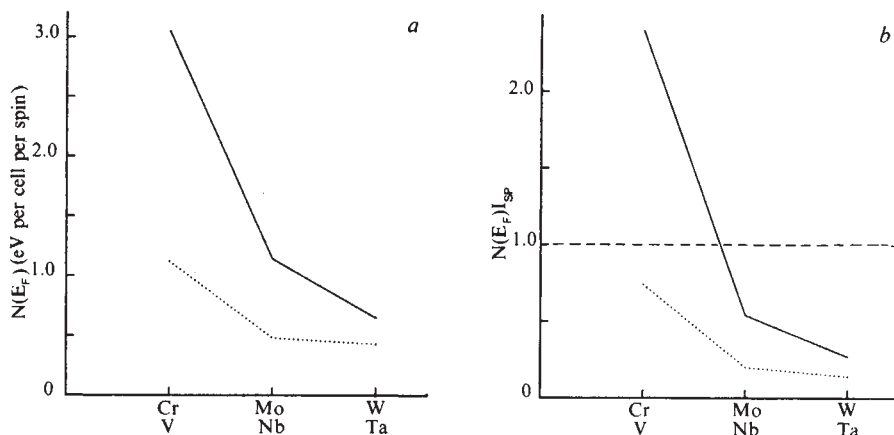


Fig.1 a, Densities of electronic states at the Fermi level, $N(E_F)$, versus metal atom in the B1-structure nitrides. **b**, Stoner parameter, $N(E_F)I_{sp}$, versus metal atom in the B1-structure nitrides. $N(E_F)I_{sp} > 1$ indicates a magnetic instability. The solid (dashed) curves connect points for the Cr(V) columns in the Periodic Table.

- Weertman, J. *Phil. Trans. R. Soc. A288*, 9 (1978).
- Anderson, D.L. *Nature* **307**, 114 (1984).
- Stacey, F.D. *Physics of the Earth* (Wiley, New York, 1969).
- Whaler, K.A. *Nature* **306**, 117 (1983).
- Galasso, F.S. *Structure, Properties and Preparation of Perovskite-type Compounds* (Pergamon, Oxford, 1969).
- Poirier, J.P., Peyronneau, J., Gesland, J.Y. & Brebec, G. *Phys. Earth planet. Inter.* **32**, 273 (1983).
- O'Keeffe, M. & Bovin, J.O. *Science* **206**, 599 (1979).
- O'Keeffe, M. in *Fast Ion Transport in Solids — Solid-State Batteries and Devices* (ed. van Gool, W.) 233 (North-Holland, Amsterdam, 1973).
- Schulz, H. A. *Rev. Mater. Sci.* **12**, 351 (1982).
- Vashishta, P., Murphy, J.N. & Sheney, G.K. (eds) *Fast Ion Transport in Solids — Electrodes and Electrolytes* (North-Holland, Amsterdam, 1979).
- Parsonage, N.G. & Staveley, L.A.K. *Disorder in Crystals*, 240, 512, 696 (Clarendon, Oxford, 1978).

should be undertaken.

We have calculated by the same method⁸ the electronic structure and the superconducting properties of VN, CrN, TaN and WN as well as NbN and MoN (Fig. 1a). In going from the group v to the group vi compounds, $N(E_F)$ increases substantially, which is a signal for superconductivity or, possibly, for magnetism. Note that MoN has a value of $N(E_F)$ more than twice that of NbN, which is itself a high- T_c material.

To quantify the tendency towards magnetism, we have plotted in Fig. 1b the quantity NI_{sp} , the product of $N(E_F)$ and the generalized exchange matrix element calculated by the Stoner theory^{9,10}. CrN, with a very large $N(E_F)$, is clearly predicted to be magnetic in the B1 structure, in agreement with experiment. VN and MoN have large values of $N(E_F)$, but nevertheless MoN does not come near the Stoner criterion ($NI_{sp} > 1$) for magnetism. It has been suggested¹¹ that in VN, spin fluctuations suppress T_c , but in spite of the large value of $N(E_F)$, MoN has a much lower value of NI_{sp} , not unusual for good superconductors. We conclude that the role of spin fluctuations in MoN is not dominant and therefore leaves it as a prime candidate for high T_c .

The strength of the electron-phonon interactions assessed by $\eta = N(E_F) I_{ep}^2$ (where I_{ep}^2 is the electron-ion scattering matrix element), reveals that both the metal and the nitrogen contributions show a very distinct maximum at MoN (Fig. 2a). It is interesting that the quantity η , unlike Stoner's $N(E_F)I_{sp}$, which scales as $N(E_F)$, does not follow $N(E_F)$. Thus CrN, despite a very large $N(E_F)$, has η smaller than that of MoN. We have also used neutron-scattering data¹² and Debye temperature results¹³ for the $NbC_{1-x}N_x$ system to estimate the force constants $M\langle\omega^2\rangle$ and hence to obtain the coupling constant λ , which reaches a pronounced maximum for MoN (Fig. 2b).

Application of the Allen-Dynes¹⁴ form of the McMillan equation gives the superconducting temperatures shown in the table. It will be seen that T_c for VN is overestimated by about a factor of two, which

Comparison of calculated and measured values of the superconducting transition temperature

	T_c^{cal}	T_c^{exp}
VN	19.7	8.6
NbN	17.1	16.0
TaN	14.6	8.2
CrN	12.4	—
MoN	29.4	—
WN	15.8	—

can be accounted for by the argument that spin fluctuations, not included in our theory, would reduce the theoretical value. The discrepancy for CrN, which is magnetic, is also explicable. Our calculation for NbN is in very good agreement with experiment¹⁵, while MoN stands out with a predicted $T_c \sim 30K$. The calculated

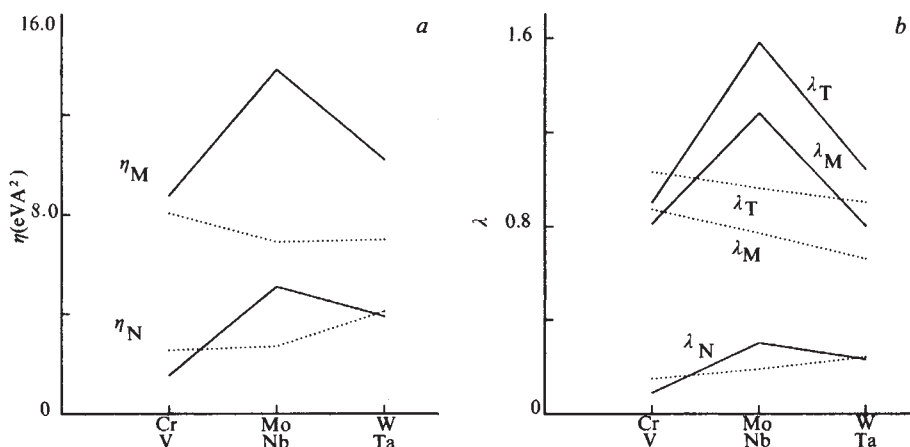


Fig. 2 a, Electron-phonon interaction parameters, η , for the nitrogen and metal sites in the B1-structure nitrides. b, Electron-phonon mass enhancement factor λ , for the B1-structure nitrides. $\lambda_{TOT} = \lambda_{MET} + \lambda_N$. The solid (dashed) curves connect points for the Cr(V) columns in the Periodic Table.

value of T_c for TaN, approximately 15K, differs from the measured value¹⁶ of about 8K, but we believe that much of this discrepancy arises because we have estimated the phonon moments by scaling from measurements of the $NbC_{1-x}N_x$ system¹³.

Transition-metal carbonitrides often form with vacancies on the non-metal sublattice, and we have estimated the effect of nitrogen vacancies on our prediction by coherent potential approximation (CPA) calculations for MoN_x . Our model, previously applied to NbC_x and TaC_x (ref. 4), assumes that the metal sublattice remains unchanged while the nitrogen sublattice contains vacancies at random. We have used the tight-binding form of the CPA, which involves a tight-binding fit to the MoN energy bands and a subsequent determination of self-energies Σ_s and Σ_p corresponding to the s and p nitrogen orbitals. Our results for $x = 0.99, 0.98, 0.95$ show that $N(E_F)$ remains within 1 per cent of the stoichiometric value. Although E_F lies at a peak in $N(E)$, it is not a narrow peak (contrary to the case in some high- T_c compounds) and so is not sensitive to disorder. So even with a modest proportion of N vacancies, MoN should have a high T_c in the B1-lattice form.

The only experimental report of B1 MoN is that of Saur and collaborators¹⁷, who identified it as a second phase in a sample that was primarily hexagonal ('WC type'). Their value of the cubic lattice constant, $a = 4.16 \text{ \AA}$, is 2 per cent smaller than our estimated value of $4.25 \text{ \AA}$ derived from measured lattice constants in the MoC_x and $NbC_{1-x}N_x$ systems. If $4.25 \text{ \AA}$ is the correct value, using the lattice constant behaviour in MoC_x as a guide indicates that a sample with $a = 4.16 \text{ \AA}$ contains 30–40 per cent N vacancies. It would be very useful to have the work of Saur *et al.* verified independently. Recent experiments by Fuller *et al.*¹⁸ using radio-frequency reactive sputtering led to the discovery of a new metastable phase of hexagonal Mo_2N , but the B1 structure was not realized.

We wish, finally, to emphasize that the

compounds which are known to form in both the B1 and hexagonal (WC) phases have the highest T_c in the B1 phase. For example, WC-structure MoC has $T_c = 9K$ compared with 14K in the B1 structure. Considering that WC-structure MoN has been reported to have T_c of 15K (ref. 19), it may be anticipated that B1-structure MoN would have considerably higher T_c . These arguments, as well as the empirical rules for the carbides and nitrides²⁰, support the idea of high T_c in B1-structure MoN. □

1. Klein, B.M. & Papaconstantopoulos, D.A. *Phys. Rev. Lett.* **32**, 1193 (1974).
2. Klein, B.M., Papaconstantopoulos, D.A. & Boyer, L.L. in *Superconductivity in d- and f-Band Metals* (ed. Douglass, D.H.) 339 (Plenum, New York, 1976); Klein, B.M. & Pickett, W.E. in *Superconductivity in d- and f-Band Metals* (eds Buckel, W. & Weber, W.) 477 (Kernforschungszentrum Karlsruhe GmbH, Karlsruhe, 1982).
3. Papaconstantopoulos, D.A. in *Physics of Transition Metals 1980* (ed. Rhodes, P.) 563 (Inst. of Phys. Conf. Ser. no. 55, London, 1981).
4. Klein, B.M., Papaconstantopoulos, D.A. & Boyer, L.L. *Phys. Rev. B* **22**, 1946 (1980).
5. Pickett, W.E. *Phys. Rev. B* **25**, 745 (1982).
6. Pickett, W.E., Klein, B.M. & Papaconstantopoulos, D.A. *Physica* **107B**, 667 (1981).
7. Jehn, H. & Eittemayer, P. *J. Less-Common Metals* **58**, 85 (1978).
8. Gaspari, G.D. & Gyorffy, B.L. *Phys. Rev. Lett.* **28**, 801 (1972).
9. Stoner, E.C. *Proc. R. Soc. A* **154**, 656 (1936); **169**, 339 (1939).
10. Wohlfarth, E.P. in *Physics of Transition Metals* (ed. Rhodes, P.) 161 (Inst. of Phys. Conf. Ser. no. 55, London, 1981).
11. Riethel, H., Winter, H. & Reichardt, W. *Phys. Rev. B* **22**, 4284 (1980).
12. Weber, W. *Phys. Rev. B* **8**, 5082 (1973).
13. Roedhammer, P., Gmelin, E., Weber, W. & Remiko, J.P. *Phys. Rev. B* **15**, 711 (1977).
14. Allen, P.B. & Dynes, R.C. *Phys. Rev. B* **12**, 905 (1975).
15. Francavilla, T.L., Wolf, S.A. & Skelton, E.F. in *Applied Superconductivity Conf. IEEE Trans. on Mag.* (1980).
16. Kilbane, F.M. & Habis, P.S. *J. Vac. Sci. Technol.* **13**, 107 (1975).
17. Saur, E.J., Schechinger, H.D. & Rienderer, L. *IEEE Trans. MAG-17*, 1029 (1981); Bauer, H., Saur, E. & Schechinger, D. *LT-14*, p.55.
18. Fuller, W.W., Wolf, S.A., Gubser, D.U., Skelton, E.F. & Francavilla, T.L. *J. Vac. Sci. Technol.* **A1**, 517 (1983).
19. Vandenberg, J.M. & Matthias, B.T. *Mat. Res. Bull.* **9**, 1085 (1974).
20. You-Xiang Zhao & Shou-an He in *Proc. 9th Int. Conf. High-Pressure Science & Technol.* (eds Whalley, E. & Homan, E.G.) (Plenum, 1983).

The authors are in the Electronic Structure of Solids Section, Condensed Matter Physics Branch, at the Naval Research Laboratory, Washington, DC 20375.

Neurobiology

Postsynaptic densities clarified?

from Gerry Shaw

THERE seems to be a curious rule concerning our understanding of elements of the neuronal cytoskeleton. The components which were discovered first have proved to be the most difficult to characterize biochemically and comprehend functionally. Thus neurofilaments and postsynaptic densities (PSDs) were visualized both at the light and electron microscope levels well before microtubules and microfilaments, yet the major subunits of microtubules and microfilaments were already characterized at a time when our knowledge of the biochemical make-up of neurofilaments and PSDs was confused. For neurofilaments, the mists began to clear about nine years ago as a result of data derived from studies of axonal transport¹. But PSDs have remained shrouded in mist. In the light of a few recent papers on the subject it is worth asking when that is likely to change.

In the central nervous system, the PSD complex is visualized by electron microscopy of ultra-thin sections as a clump of very electron-dense material apparently adherent to the postsynaptic membrane. Fine filaments can often be seen to extend from this dense clump. The size of the complex is variable, as is the shape, and it has been suggested that such morphological differences are correlated with neurotransmitter type². Much work has been directed towards identifying the protein constituents of PSDs. Presumably the protein that makes the fine filaments is a major component of total PSD protein. One would also expect to find some sort of protein or glycoprotein involved in gluing the complex to the postsynaptic membrane. Finally, if the complex acts as an anchoring structure, it is possible that a variety of different synapse-specific enzymes and receptors might be bound to it. These constituents would add biochemical complexity to the PSD complex. A further complication is that different proteins may be associated with different types of PSD. Accordingly, it is perhaps not surprising that the PSD has been so intractable.

An obvious way to find out which proteins make up PSDs is to examine biochemically a fraction of brain material which, by electron microscope criteria, is composed predominantly of PSDs. Using this approach, actin, fodrin, tubulin, calmodulin, synapsin I and a 50,000-molecular-weight (50K) protein (called the 'major postsynaptic density protein') have all been postulated as components of the PSD. Another approach has been to react sections of brain tissue with antibodies specific for various known proteins and to see, by light and electron microscopy, whether PSDs

are stained. Actin, tubulin, an intermediate filament protein, synapsin I, calmodulin and microtubule-associated protein 2 have been implicated as part of the PSD by this approach. In addition, a group of glycoproteins found in synaptic junction preparations and originally identified by their ability to bind lectins, could be PSD constituents (see, for example, ref. 3).

If all these results are correct, the PSD must be an extremely complex structure. So could some of them be due to co-purification of proteins found outside the PSD complex or to spurious immunoreactivities? Some data included in two recent papers, suggest that could sometimes be the case.

Thus Matus *et al.* found neurofilament proteins and glial fibrillary acidic protein (GFA), the components of the intermediate filaments of neurones and astrocytes, in the purified PSD fraction routinely used for biochemical experiments⁴ and went on to show that purified radiolabelled GFA, added to brain material at the start of the standard PSD preparation, also ended up in the final PSD fraction⁴, clearly indicating that its presence there was artefactual. Therefore the presence of a protein in the standard PSD preparation does not necessarily indicate that this protein is also an *in vivo* component. And a carefully controlled immunoelectron microscopical study of synapsin I showed that claims it was a true PSD constituent were wrong and probably founded on non-specific binding of the antibody to PSDs⁵. So immunological approaches may also be problematic; the PSD seems to be a 'sticky' structure presenting special problems to the biochemist and the immunologist.

More recently, Landis and Reese⁶ examined the structure of dendritic spines in the mouse cerebellum using John Heuser's technique of rapid freezing, deep etching and rotary shadowing. They found that the PSD was composed of a mass of fine filaments with a diameter of about 4 nm, much smaller than that of actin-containing microfilaments clearly visualized in other regions of the dendritic spine. The diameter also makes it unlikely that the PSD filaments are composed of microtubule or intermediate filament proteins, at least in their usual conformations. What, then, are these filaments made of?

It would have been neat if the answer had been the 50K protein of PSD. However, the protein is apparently absent from cerebellar PSDs and, in an important paper, Kennedy *et al.* recently demonstrated that the 50K protein is indistinguishable from a component of a calmodulin-dependent protein kinase of brain tissue⁷ and is part of

an abundant 650,000-molecular-weight holoenzyme⁷, data since confirmed by Kelly *et al.*⁸ Moreover, Kennedy's group suggests that the complex is composed of nine α (50K) subunits and three β (60K) subunits, both of which can be detected in PSD preparations⁹. Gel filtration studies suggest that the diameter of the holoenzyme should be about 20 nm, and it is interesting that particles of about this size have been detected in electron microscopical studies of PSD preparations¹⁰. Since Kennedy *et al.* have a monoclonal antibody against the α subunit they should be able to check whether the protein is a true component of PSDs *in vivo*. In any case, it seems inconceivable that the PSD filaments are composed of subunits of the enzyme.

Bearing these results in mind, what do we really know about the components of PSDs? The answer seems to be that we can still only guess at the make-up of the PSD filaments, the dense material and the membrane attachment, if these are indeed distinct elements. We should, at least, soon know whether the major PSD protein is really a component of the PSD complex *in vivo* and, if so, in what form it appears in the complex. For the future, comparison of PSDs with some of the better understood cell-to-cell and cell-to-substrate contacts might also be illuminating. We might find proteins in PSDs related to proteins of desmosomes or adhesion plaques. Perhaps we must look for the protein components of the PSD filaments in the many minor bands found on polyacrylamide gels of PSD preparations (see ref. 11 for example).

It is even possible that the filament component may have been completely missed due to its lack of solubility in gel sample buffer, as was the case with the protein or proteins of the paired helical filaments found in the brain of patients suffering from senile dementia of the Alzheimer's type¹². It has been supposed that such paired helical filaments are composed of modified neurofilament proteins, though recent data do not confirm this view, and indeed leave the identity of the subunits of these filaments an open question¹². Since the paired helical filaments are found in neurones which typically have thousands of PSDs, could it be that they are the result of overproduction of a PSD filament protein? At present there are not enough data to muzzle speculation. □

- Hoffman, P. N. & Lasek, R. J. *J. Cell Biol.* **66**, 351 (1975).
- Gray, E. G. *J. Anat.* **93**, 420 (1959).
- Mena, E. E. & Cotman, C. W. *Science* **216**, 422 (1982).
- Matus, A. *et al.* *J. Cell Biol.* **87**, 346 (1980).
- De Camilli, P. *et al.* *J. Cell Biol.* **96**, 1355 (1983).
- Landis, D. & Reese, T. *J. Cell Biol.* **97**, 1169 (1983).
- Kennedy, M. B. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 7357 (1983).
- Kelly, P. T., McGuinness, T. L. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 948 (1984).
- Bennet, M. K. *et al.* *J. Biol. Chem.* **258**, 12735 (1983).
- Blomberg, F. *et al.* *J. Cell Biol.* **74**, 204 (1977).
- Matus, A. *et al.* *J. Cell Biol.* **98**, 237 (1984).
- Ihara, Y. *et al.* *Nature* **304**, 727 (1983).

Gerry Shaw is at the Max Planck Institute for Biophysical Chemistry, D-34 Goettingen, FRG.

Geophysics

Crustal thickening in Himalayas and Caledonides

from Drummond Matthews and Alfred Hirn

WE wish to point out some striking similarities that have been revealed by recent geophysical investigations^{1,2} of two collision mountain belts, the major orogenic belts of the Himalayas^{3,4} and the Caledonides⁵.

The spectacular mountains of the Himalayas are the result of collision between India and Eurasia, and gravity measurements⁶ indicate a doubling in thickness of the continental crust beneath both the High Himalayas and the Tibetan plateau. This has usually been regarded as having occurred by the sliding of the entire Indian crust under that of Eurasia, but explosion seismology by a joint French-Chinese team¹ suggests a more complex interdigitation of slices of lower and upper crusts. About 400 million years ago, a similar collision between continents was probably responsible for the creation of the Caledonide belt, whose dismembered roots can now be traced from Florida to East Greenland and northern Norway. Results from the BIRPS reflection profiling project² indicate a gross crustal structure remarkably comparable to that of the Himalayas.

The French-Chinese expedition shot a long reversed critical to wide-angle seismic profile in southern Tibet, parallel to the trend of the High Himalayas. This established the crustal velocity structure and found the depth of the Moho to be 70 km. The expedition also made a fan profile along a road running across the Himalayas into Nepal, recording super-critical reflections from the Moho. These arrivals are easily seen on the seismograms and, assuming the velocity structure observed north of the Himalayas, have been converted to depth to provide the basis for Fig. 1a, with the Moho as a marker horizon. The figure suggests that crustal thickening takes place by large-scale imbrication: imbrication of the upper crust occurs along (mapped) northerly dipping thrusts, the Main Boundary Thrust and the active Main Central Thrust, and imbrication of slices of the lower crust and uppermost mantle along (inferred) southerly dipping thrusts.

The BIRPS deep seismic reflection profile extends from the north coast of Scotland across the Lewisian foreland west of the Hebrides, passes between Ireland and Scotland and terminates off central Wales. It crosses all the major faults of the British Caledonides. For most of the 1,000 km long line, the lower crust is strongly reflective, characterized on unmigrated sections by short (<3 km) discontinuous reflections. The Moho, labelled by refraction experiments, is believed to

coincide with the bottom of this reflection zone⁷.

Major thrusts known from surface mapping on land, notably faults such as the Sgurr Beag slide, the Moine Thrust, the Outer Isles Thrust and the South Irish Sea lineament, can be traced down to the reflective lower crust where they are lost. The strongest reflector in the upper mantle has been dubbed the 'Flannan Thrust'. Naturally, the implied interpretation is speculative, but the feature appears to be subparallel to the Outer Isles Thrust. A similar reflector appears off northern Ireland where we believe it to be displaced by the transcurrent Great Glen Fault. The Flannan Thrust can be seen as a very strong reflector coming up from the base of the section at about 45 km (Fig. 1b). It is lost in the layered lower crust. We suggest that this feature is similar to those southerly dipping thrusts inferred from the Himalayan cross-section, which bring up a slice of lower crust, Moho and uppermost mantle.

We believe that imbrication of the continental lithosphere occurs in Tibet along thrust faults both near the top surface and at the interface between the crust and mantle, and that these depths correspond to brittle layers in which earthquakes originate. Chen and Molnar⁸ have demonstrated that earthquakes occur in Tibet in

two depth ranges: 5–10 km in the upper crust and 85–90 km just beneath the Moho. Following Meissner⁹, they observe that these depths correspond to the two strong but brittle zones in a lithosphere whose rheology depends primarily on temperature. We may expect the quartzo-felspathic crust to be cool and brittle at the top, warmer and more ductile at the bottom. The olivine of the upper mantle will be brittle at the temperature of the crust-mantle interface but become ductile where it is deeper and hotter.

Arguments about the origin of the reflections from the lower crust that have now been observed on normal incidence profiles in three continents turn on the effects of free water or of ductile deformation of the rocks^{10–13}. It is tempting to speculate that the reflective lower crust is also the ductile lower crust and it would be of interest to obtain a profile across an active thrust zone, where earthquake depths and mechanisms had been determined.

Thrust faults that are no longer seismically active — the Wind River Thrust in Wyoming¹⁴ and the Outer Isles Thrust, for example are known to be associated with reflectors, although the physics of the reflective zone is not understood. Reflections from these thrust fault zones have been traced down into the lower crust, where they are lost above the Moho. Likewise, the Flannan Thrust is a strong reflector in the uppermost mantle but is lost in the layered ductile lower crust. We believe that these features, seen on profiles of the 400 million year old Caledonides, are analogues of the present-day thrusts inferred beneath the Himalayas. If that is so, then crustal thickening during the Caledonian orogeny may also have

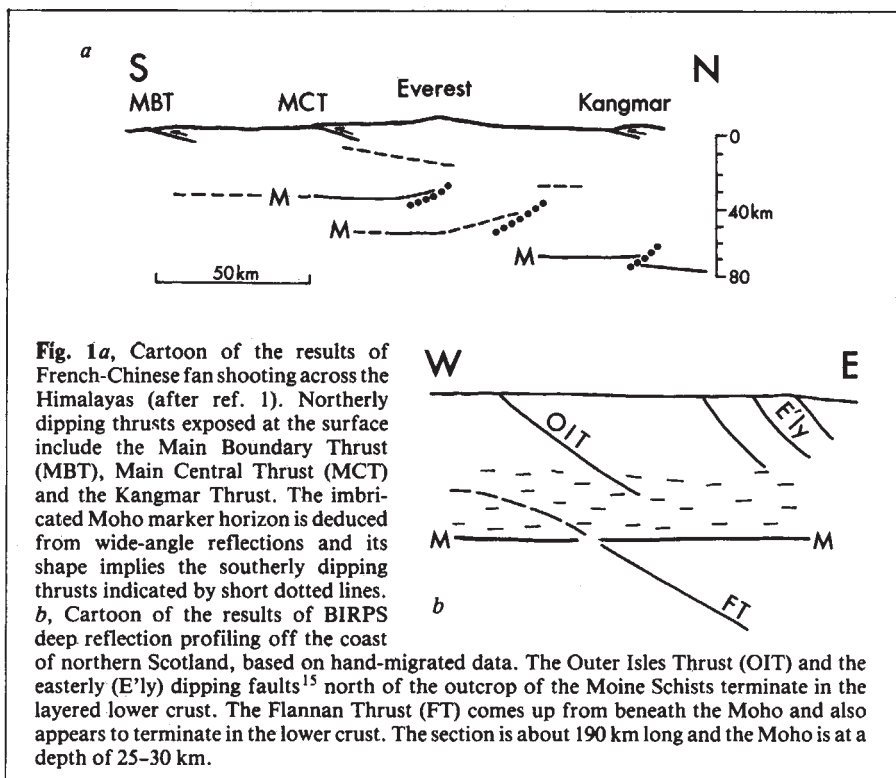


Fig. 1a, Cartoon of the results of French-Chinese fan shooting across the Himalayas (after ref. 1). Northerly dipping thrusts exposed at the surface include the Main Boundary Thrust (MBT), Main Central Thrust (MCT) and the Kangmar Thrust. The imbricated Moho marker horizon is deduced from wide-angle reflections and its shape implies the southerly dipping thrusts indicated by short dotted lines. **b,** Cartoon of the results of BIRPS deep reflection profiling off the coast of northern Scotland, based on hand-migrated data. The Outer Isles Thrust (OIT) and the easterly (E'ly) dipping faults¹⁵ north of the outcrop of the Moine Schists terminate in the layered lower crust. The Flannan Thrust (FT) comes up from beneath the Moho and also appears to terminate in the lower crust. The section is about 190 km long and the Moho is at a depth of 25–30 km.

occurred "not as the result of the superposition of two whole normal crusts on each other, but rather as the result of separate doubling (by decoupling and thrusting) of the upper crustal and lower crustal layers (plus Moho)"¹. Whether this model can be applied to other orogenic belts remains to be seen. □

1. Hirn, A. *et al.* *Nature* **307**, 23 (1984).
2. Brewer, J.A. *et al.* *Nature* **305**, 206 (1983).
3. Dewey, J.F. & Bird, J.M. *J. geophys. Res.* **75**, 2625 (1970).
4. Powell, C.McA. & Conaghan, P.J. *Earth planet. Sci. Lett.* **20**, 1 (1973).
5. Watson, J. & Dunning F.W. in *The Caledonides of the British Isles Reviewed* (eds Harris, A.L., Holland, C.H. & Leake, B.E.) 67 (IGCP Proj. 27, Scottish Academic Press, 1979).
6. Lyon-Caen, H. & Molnar, P. *J. geophys. Res.* **88**, 8171 (1983).
7. Barton, P., Matthews, D.H., Hall, J. & Warner, M.R. *Nature* **308**, 55 (1984).
8. Chen, W.P. & Molnar, P. *J. geophys. Res.* **88**, 4183 (1983).
9. Meissner, R. *J. Geophys.* **40**, 57 (1974).
10. Oliver, J., Cook, F. & Brown, L. *J. geophys. Res.* **88**, 3329 (1983).
11. Mathur, S.P. *First Break* **1**(7), 9 (1983).
12. Hale, L.D. & Thompson, G.A. *J. geophys. Res.* **87**, 4625 (1982).
13. Fuchs, K. *Zeitschr. Geophys.* **35**, 133 (1969).
14. Smithson, S.B., Brewer, J.A., Kaufman, S. & Oliver, J.E. *J. geophys. Res.* **84**, 5955 (1979).
15. Brewer, J.A. & Smythe, D.K. *J. geol. Soc.* **141**, 105 (1984).

Drummond Matthews is at the Department of Earth Sciences, Bullard Laboratories, University of Cambridge, Cambridge CB3 0EZ, and Alfred Hirn at the Institut de Physique du Globe, Université Pierre et Marie Curie, 75230 Paris Cedex 05.

Crop science

The genetic manipulation of crop plants

from Ben Mifflin and Peter J. Lea

TWO recent UK symposia* devoted to the genetic manipulation of plants addressed the same question: 'what has it done for agriculture?'. The superficial answer might be: 'so far, not much'. But that would be to misunderstand the long-term nature of crop science. Bear in mind, for example, that although the initial sexual cross in the production of the currently successful wheat cultivar Avalon was made in 1969, the variety was only ready to be released in 1980 — a period that more than encompasses the whole history of recombinant DNA technology. Moreover, the amount of information available and the number of systems under study are already impressive compared with what was known and foreseen five years ago. What then has been achieved?

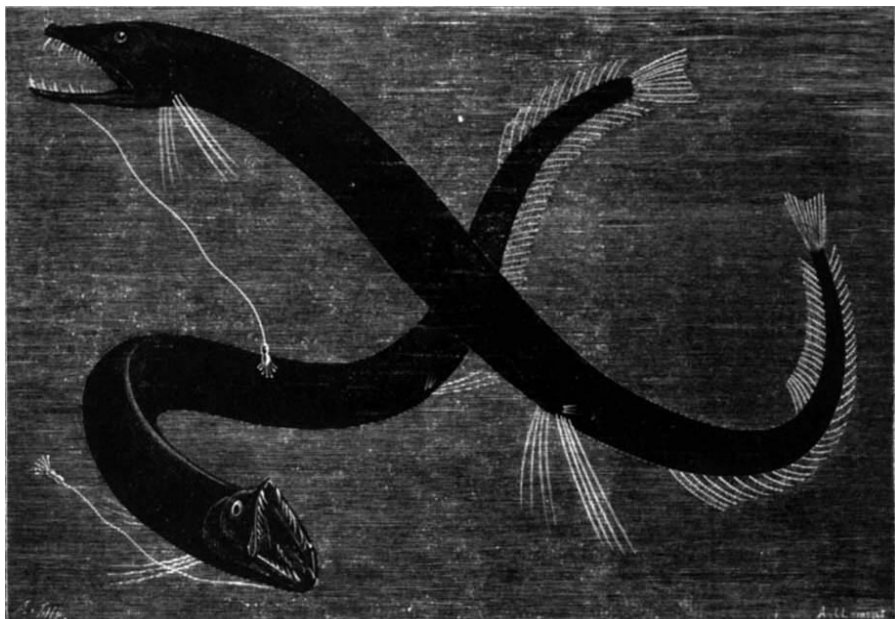
For a start, recombinant DNA techniques for gene isolation have been applied to a wide range of crop species. The work

has centred on the genes for the most abundant proteins, such as ribulose biphosphate carboxylase, an enzyme of the carbon-fixation cycle, and the seed storage proteins of cereals and legumes. From this, a pattern of organization is emerging in which each major group of cereal proteins is encoded by one or more complex multi-genic loci, the individual genes of which contain a number of different tandem and interspersed repeats (B. Forde *et al.*, Rothamsted). The genes themselves, particularly those of the legume storage proteins, appear to have a conventional structure of coding sequences interspersed with introns (D. Boulter, University of Durham). Currently the flanking regions of various plant genes are being searched and tested for sequences that are important in controlling the expression of genes particularly in response to environmental stimuli such as stress or light.

Although the study of plant protoplasts is expanding and the list of species that can be regenerated from protoplasts grows ever larger, monocotyledonous species are notable by their absence and techniques producing large number of regenerants at high frequency have only been worked out for very few species. Protoplasts are potentially valuable for at least three reasons. First, new genes can be introduced into plants through protoplasts, whose co-cultivation with *Agrobacterium* appears to be an efficient way of going about the task (J. Schell *et al.*, Max-Planck-Institut, Köln). Second, the selection of plant mutants by screening protoplasts is, in theory, valuable. It has, for example, led to mutants deficient in nitrate reductase (R. Mendel and A. Muller, Akademie der Wissenschaften, Gatersleben, GDR), but in general the technique has been somewhat disappointing and many workers have turned to other methods. Third, protoplast fusion offers a way of combining species that are not sexually compatible.

Although progress in the development of protoplast and tissue culture techniques has been slow and frustrating, the techniques have an immediate application. Micropropagation of high-value crops is of increasing use in the horticultural industry (G. Hussey, John Innes Institute) and there is considerable interest in the exploitation of the variation ('somaclonal variation') induced by passage of plants through

*ARS genetic manipulation of crop plants: 5 years on', organized by the Agriculture and Food Research Council, Cambridge, 12–13 December 1983, and 'The genetic manipulation of plants and its application to agriculture', organized by the Phytochemical Society of Europe, London, 19–20 December 1983.



100 years ago

DEEP-SEA FISHES OF THE "TALISMAN"

IN THE exhibition, now open at the Jardin des Plantes, Paris, the great interest of the fish captures of the *Talisman* centres in the remarkable forms taken from the depths of the sea. The question of whether certain fish inhabit certain zones of depths was closely considered, and is answered in the affirmative. These zones are of very considerable depth, varying from 600 to over 3650 metres, and in bringing up specimens from such areas of great pressure these suffer immensely through the phenomena caused by the rapid decompression of the air, the more remarkable effects being dilation of the swim bladder, the eyes being squeezed out of

their orbits, and the scales clothing the body are shed. In some cases even the fishes body has become smashed into pieces.

In some of the deep-sea fishes peculiar organs, unknown for the most part among surface fishes are to be found. Long filamentous organs are to be met with showing apparently a brilliant type of phosphorescence. One of the most singular is that to be seen on a fish represented in the annexed woodcut. In this form (*Eustomias obscurus*) the tactile organ takes the appearance of a long filament, which is placed underneath the lower jaw, and which ends in an inflated and rayed knob-like phosphorescent mass.

From *Nature* **29**, 483, 20 March 1884.

protoplast or tissue culture and the use of tissue cultures in the manipulation of ploidy (D. Ingram *et al.*, University of Cambridge). Hundreds of lines of potato and wheat plants derived from culture are being evaluated by state and commercial breeders.

The application of precise gene transfer to yield new cultivars seems some way off, even though crop plants such as potato have been successfully transformed and taken through a tuber generation (G. Ooms *et al.*, Rothamsted). The reason is the absence of genes, and even of the knowledge of which genes, to transfer. Since this problem stems largely from gaps in understanding of plant biochemistry, many of the techniques described above are being used to investigate aspects of plant metabolism.

Nitrogen and carbon metabolism have been the first to be attacked. The contribution of the *Rhizobium* genome to nitrogen fixation in the root nodules of legumes is being analysed rapidly and genes for fixation, nodulation and symbiosis have been characterized (J. Downie and A. Johnston, John Innes Institute, and see these columns in *Nature* 306, 639; 1984). Plant genes involved in the symbiosis that have been isolated include that for leghaemoglobin (D. Collinge *et al.*, Aarhus University) and, more recently, glutamine synthetase (J. Cullimore *et al.*, Rothamsted). Cloning of other genes important in nitrate and ammonia assimilation is reasonably advanced only in prokaryotes (J. Wootton, University of Leeds). Genes for both subunits of ribulose biphosphate carboxylase have been isolated; since that for the large subunit can be expressed in *Escherichia coli*, the way is open to explore structure-function relationships through *in vitro* mutagenesis (R. Ellis, University of Warwick, and A. Gatenby, Plant Breeding Institute). Directed *in vitro* mutagenesis affecting the active site of the single-subunit carboxylase of *Rhodospirillum rubrum* has already been achieved (S. Gutteridge and G. Lorimer, Du Pont de Nemours, Wilmington). Nitrogen and carbon metabolism have also been investigated with defined biochemical mutants. This has led to a better understanding of the pathways and their regulation as well as to new lines of barley with greater amounts of nutritionally essential amino acids in the seed (S. Bright *et al.*, Rothamsted).

Another active area of molecular biological investigation is the level at which genes are controlled in the defence mechanisms of plants against disease and stress as well as in the normal development of plants. For example, considerable progress has been made in identifying mRNAs involved in the early steps of phytoalexin (plant fungicides) biosynthesis (K. Hahlbrock *et al.*, Max-Planck-Institut, Köln), and the heat-shock proteins of plants and their mRNAs have been identified (F. Schoffl *et al.*, Bielefeld, FRG and

Athens, Georgia). In both systems the synthesis of mRNAs appears to be largely controlled at the level of transcription. Other systems, yielding equivalent information, include the induction, by light, of the genes of certain chloroplast proteins (T. Gallagher *et al.*, University of Warwick) and the activation of genes involved in fruit ripening (D. Grierson *et al.*, University of Nottingham).

In summary, despite the activity and considerable progress reported, the genetic manipulation of crop plants is still only getting underway. For the present,

improvements in agricultural plants will continue to come from that older and well-established technique of genetic manipulation, plant breeding. This technology (as described by J. Bingham, Plant Breeding Institute) is highly sophisticated; it can safely be predicted that the earliest successes of the recombinant DNA approach will enhance that sophistication rather than replace it. □

Ben Mifflin and Peter J. Lea are members of the Biochemistry Department, Rothamsted Experimental Station, Harpenden, Herts AL5 2JQ.

Discovery of a comet

IN October 1892, when the American astronomer E.E. Barnard took the dry plate shown below, comet photography was not novel. As early as 1858, Donati's Comet was recorded on wet collodion plates by an English photographer named Usherwood. Sir David Gill obtained remarkably fine dry plates of the Great September Comet (1882 II), and cometary spectra were successfully photographed the previous year. But the image below marks the first photographic discovery of a comet.

It was a measure of the increasing role of photography in studying celestial bodies that only a month earlier Barnard had made what proved to be the last visual discovery of a satellite (Jupiter V, or Amalthea). In the course of his monumental survey of the northern Milky Way, Barnard photographed a field in the constellation Aquila with the 15-cm Willard portrait lens at Lick Observatory, California. When the plate was developed, a comet (1892 V) — or rather the trailed image produced by the comet's motion over four hours with respect to the

background of distant stars — serendipitously appeared dead centre. Barnard had failed to spot it visually despite continuous observation through the guiding eyepiece during the long (260 min) exposure. As he later wrote, "Nothing whatever was known of its existence until it was shown on the photographic plate. It was then looked up in the sky and observed with the micrometer, i.e. visually through a telescope. . . Its history is complete, and until better evidence is forthcoming it will stand as the first comet whose original discovery was made by photography".

With a telescope of larger aperture, visual corroboration was immediately forthcoming. Photographs taken on succeeding nights were used to determine the orbit, recalculated by D. Yeomans in 1974 to yield a perihelion distance of 1.43 AU and a period of 6.52 years. **Jon Darius**

The photograph is one in an exhibition 'Beyond Vision' at the Science Museum, London from 19 April, and is reproduced here by courtesy of the Royal Astronomical Society. A book of the same title will be published in April by Oxford University Press.



Metastases limited

SIR — It has recently been reported by Kun *et al.*¹ that poly (ADP-ribose) has an important role in carcinogenesis, and that inhibition of the biosynthesis of this polymer from NAD inhibits carcinogenesis^{2,3}. Based on the structural analogy with nicotinamide, these authors tested benzamide and found it to inhibit this biosynthesis, and to prevent carcinogenesis in human fibroblasts².

In Padua we have conducted studies of anticancer activity of analogues of nicotinamide, aiming at interfering with NAD activities³. One promising compound, 6,6'-dithiodinicotinic acid (carboxypyridine disulphide, CPDS) was selected for further studies, also because of its unusual reactivity properties. CPDS was found to decrease by 75% the formation of lung metastases in mice bearing Ehrlich ascites carcinoma⁴; later studies showed CPDS to reduce by as much as 84% the metastatic spread of Lewis lung carcinoma in mice, and to significantly decrease the rate of growth of the primary carcinoma in these mice⁵. Other analogues of nicotinamide tested had similar effects⁵.

The fact that the same type of compound is active against both carcinogenesis and metastasis suggests an intimate link between the two. Preliminary experiments indeed appear to confirm this relationship.

D. R. GRASSETTI

Institute of Pharmacology,
University of Padua,
35100 Padua, Italy

1. Kun, E., Kirsten, E., Milo, G.E., Kurian, P. & Kumari, H.L. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7219-7223 (1983).
2. Kun, E. *et al.* in *Twelfth Steenbock-Lilly Symposium* (eds Lennon, D.L.F., Stratman, F.W. & Zahiten, R.N.) 387-399 (Elsevier, New York, 1982).
3. Grasseti, D.R. *Chemtech* 666-671 (November, 1973).
4. Grasseti, D.R. *Nature* **228**, 282-283 (1970).
5. Grasseti, D.R. in *Current Chemotherapy and Immunotherapy, Proc. 12th Int. Congress Chemotherapy, Florence Vol II*, 1365-1366 (American Society of Microbiology, Washington, D.C., 1981).

Flagella or cilia?

SIR — Margulis¹ emphasized the morphological case for separate terms to denote "flagella" of prokaryotes and eukaryotes. But the proposed resurrection of "undulipodia" for those of the latter (and cilia) is objectionable, as it is far from euphonious and a linguistic hybrid (Latin-Greek). If a new term is required, I tentatively suggest *mastigion* (Greek: "little whip", a diminutive of *mastix*, "whip"²) — for one or other group of organelles. By priority of usage, flagellum should presumably be retained for the eukaryotic organelle, and mastigion used for that of prokaryotes. But mastigion is already familiar to protistologists from taxonomic usage (*Mastigophora* and so on), so less disruption might result from its application to the eukaryotic structure. Retention of the name flagellum in bacteria would also accord with naming their protein constituent "flagellin".

But is a new term required at all? Both eukaryotic "flagella" and "cilia" have a similar basic "9 + 2" structure; although "flagella" are usually longer and less numerous than "cilia", usually beat independently rather than in metachronal rhythm, and sometimes possess a paraxial rod. Professor K. Vickerman has pointed out (personal communication) that none of these features constitutes a constant difference: the numerous "flagella" of *Trichonympha* and *Opalina* exhibit metachrony, by no means all "flagella" have paraxial rods, and length is not an absolute criterion. There seems therefore no justification for their distinction. An alternative solution to the introduction of a new term might be to restrict "flagellum" to the prokaryotic organelle, and extend "cilium" to all the eukaryotic structures. Although the taxonomic associations of cilium make this unlikely to appeal to protistologists, metazoan cytologists would probably prefer to retain the term "cilium" than to adopt a totally new term; and objections from protistologists could be met by redefining the Ciliophora by their possession of an infraciliature rather than (less universally) cilia³. J. R. BAKER
*Institute of Terrestrial Ecology,
Culture Centre of Algae & Protozoa,
Cambridge CB3 0DT, UK*

1. Margulis, L. *Biosystems* **12**, 105-108 (1980).
2. Kenneth, J.H. in *A Dictionary of Biological Terms* 8th edn (Oliver & Boyd, Edinburgh, 1968).
3. Levine, N.D. *et al.* *J. Protozool.* **27**, 37-58 (1980).

Sea cow discovery

SIR — The recent discovery of "a virtually complete skeleton" of Steller's sea cow on Bering Island (*Nature* **306**, 415; 1983) has potential importance that may be overlooked. Although numerous skeletons (considerably more than six) of *Hydrodamalis gigas* (= "*Rhytina stelleri*") have been displayed over the past century, not one has been complete and only one (a juvenile specimen in Helsinki) is apparently associated; all the rest are composites of isolated bones. Therefore we remain largely ignorant of such details as the limb proportions, number of ribs and vertebrae, and morphology of the hand skeleton (of which not a single scrap has ever been discovered).

The fact is that several fossil sirenians, including some species directly ancestral to *H. gigas*, are better known osteologically than this one which persisted into historic times. It is to be hoped that the new specimen, if truly associated, will not gather dust in a provincial museum, but will be studied by Soviet palaeontologists and will add to our still-faulty knowledge of one of the most bizarre of all mammals.

DARYL P. DOMNING

Department of Anatomy,
College of Medicine,
Howard University,
Washington, D.C. 20059, USA

Meteorite ages

SIR — Jon Darius (*Nature* **308**, 18; 1984) says the recovery of the Lost City meteorite was the first time that photographs from synchronized cameras were used to predict the place of fall of a meteorite. It was not; the Příbram, Czechoslovakia, fall of 7 April 1959 was photographed as a bright fireball and a search of the predicted area yielded four fragments, of total weight 5.555 kg, of a possible nineteen¹. Like Lost City and the third meteorite to be recovered by photographic prediction (Innisfree, Alberta), Příbram had the orbit of an Apollo asteroid².

Darius also says that the Lost City meteorite has an age in the range 4.8–5.5 × 10⁹ years but this is older than the "age of the Solar System", about 4.55 × 10⁹ years, based on isotopic studies in meteorites. The only material claimed to be older occurs as inclusions in the Allende meteorite, two of which have yielded ages of about 4.9 × 10⁹ years. However, the interpretation of the Ar-Ar data that produced these data has been questioned (see references in ref. 3).

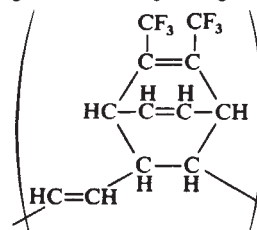
ROBERT HUTCHINSON

Department of Mineralogy,
British Museum (Natural History),
Cromwell Road, London SW7 5BD, UK

1. Rost, R. *Casopis min. geol.* **5**, 423 (1960).
2. Halliday, I. *et al.* *J. R. Astron. Soc. Can.* **72**, 15 (1978).
3. Jessberger, E.K. *Meteoritics* **17**, 232 (1982).

Unnameable named

SIR — In the report on the preparation of soluble conducting polymers (*Nature* **304**, 487; 1983) mention is made of the "unnameable" precursor polymer III with the following structural repeating unit:



In fact, the above polymer should be named following the recommendations of the International Union of Pure and Applied Chemistry¹ as poly[[5,6-bis(trifluoromethyl)bicyclo[2.2.2]octa-5,7-dien-2,3-ylene]vinylene]. The corresponding current *Chemical Abstracts* index name² is poly[[5,6-bis(trifluoromethyl)bicyclo[2.2.2]octa-5,7-diene-2,3-diyl]-1,2-ethenediyl].

KURT L. LOENING

International Committee on
Nomenclature and Symbols,
International Union of
Pure and Applied Chemistry,
PO Box 3012, Columbus,
Ohio 43210, USA

1. IUPAC, Macromolecular Division. Commission on Macromolecular Nomenclature *Pure appl. Chem.* **48**, 373-385 (1976).
2. *Index Guide* Section 222. (Chemical Abstracts Service, Columbus, Ohio, 1982).

The evolutionary relationships of man and orang-utans

Jeffrey H. Schwartz

Department of Anthropology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

Man shares uniquely few morphological features with either the chimpanzee or the gorilla, whereas there are many features that suggest affinities between man and the orang-utan, to whom the fossil Sivapithecus appears to be closely related. If these are unique features, inclusion of Sivapithecus, man and the orang-utan in a single clade, distinct from that containing the African apes, is justified but contrary to current opinion.

THERE had, until recently, been general agreement that *Homo* and *Australopithecus* (undeniable hominids) were most closely related to the African apes, among extant primates, and to the fossils *Ramapithecus*, *Sivapithecus* and *Gigantopithecus*. These fossil specimens possess, in common with hominids, low-cusped premolars and molars and thick molar enamel¹⁻⁷. But, what also emerged was the unexpected discovery that the Asian ape, the orang-utan (*Pongo*), but not the African apes, the chimpanzee (*Pan*) and the gorilla (*Gorilla*), possessed the low-cusped cheekteeth and thick molar enamel otherwise considered diagnostic of hominids. In the light of these findings Smith and Pilbeam⁸, for example, offered an ecologically based argument to explain how the distantly related hominoids *Homo* and *Pongo* could have evolved convergently similar cheektooth configurations. But, more recently, a reappraisal has been forced by the surprising discovery of new specimens of *Sivapithecus* which had the facial details of *Pongo*⁹⁻¹². Consequently, *Sivapithecus* (along with *Ramapithecus* and *Gigantopithecus*) was relegated to a separate orang-utan clade (see, for example, ref. 10). An alternative consideration is that these hominid dental features were possessed by a primate that was ancestral to *Homo*, the African apes, *Pongo* and *Sivapithecus* but were then lost during the subsequent evolution of the African apes¹³⁻¹⁵. To try to clarify the problems raised for schemes of hominoid as well as hominid evolution by recent data, I review here the morphological and other criteria on which current theories are based.

One of the most common notions about the effects of evolution is that the more closely related species are, the more similar they will be overall. In more recent years, this generalization has been considered unreliable and less than accurate. Much greater attention has been paid to the discrimination of: (1) characters that are shared by taxa (for example, species) because of the recentness of their common ancestry, from (2) those characters that are inherited as retentions from more distant, less immediately related, ancestors. It is, therefore, not necessarily the quantity of characters that indicates closeness of relatedness; rather, it is the number of features possessed in common that, because of their comparative novelty, reflect relatedness, that is, recentness of ancestry. When there are, as is frequently the case, alternative theories of relatedness, the most highly corroborated (that is, the one based on the greatest number of evolutionary novelties) would appear to be the most appropriate. In the following sections, I have tried to isolate those characters that are restricted to, and shared uniquely by, humans and another primate, or at most a few primates, in the hope of identifying that primate, or those primates, to whom we are the most closely related.

Morphology

Andrews and Cronin¹⁰ recently reviewed a series of craniofacial characters that are found in humans as well as the African apes. They pointed out (p. 543) that "many of the characters shared

by man, chimpanzees and gorillas are also present in the gibbons", which implies that these characters are "primitive for Hominoidea (man and apes) both because they are thus present in most extant hominoids and because they are found in the most widely separated members of the group".

Table 1 presents the characters studied by Andrews and Cronin but as they are distributed more broadly among extant anthropoids (hominoids, Old World monkeys (cercopithecines and colobines), New World monkeys (ceboids)) and strepsirrhines (lemurs and lorises). The enigmatic tarsier was not included because its orbital region is extremely modified. Characters cited by Andrews and Cronin as being excessively variable (for example, glabellar development) were not considered further. As the five hominoid genera, *Homo*, *Pan*, *Gorilla*, *Pongo* and *Hylobates* (the gibbon), represent only about 10% of the total sample, characters that are found exclusively in all five or fewer hominoids are certainly restricted in their distribution.

Not surprisingly, the data in Table 1 do not group man and African apes, or man and one of the African apes. The broader comparison does, however, demonstrate that extant hominoids are distinguished among primates by, for example, truncation of the subnasal plane and in having infraorbital foramina that are well removed from the zygomaticomaxillary suture. These may be added to the postcranial characters that are commonly cited as uniting extant hominoids (for example, refs 12, 16). Other characters emerge as being more restricted among extant hominoids in their expression: that is, inconsistent development of a stepped-down subnasal plane, nasal aperture shape, as well as diminution and midline appression of the anterior palatine fenestrae. These characters are found only in humans and the 'great' apes and not in gibbons and therefore seem to attest to the phylogenetic unity of the large hominoids within the hominoid clade.

Smaller groupings within the clade of large hominoids are suggested. *Homo* and *Pongo* are set apart by their low-cusped cheekteeth and thick molar enamel¹⁻⁷ as well as by the restriction of the double incisive foramina to a single opening palatally¹⁷. The African apes are themselves distinguished from other hominoids by anatomical configurations specifically related to their form of knuckle-walking^{10,16}.

In another study, Delson and Andrews¹⁸ compared numerous dental, cranial and postcranial characters among extant and fossil catarrhines (hominoids and Old World monkeys), but no feature seemed to be shared exclusively by *Homo* and the African apes. Review of their data reveals that two characters, which they interpret as derived among catarrhines, are shared uniquely by *Homo* and *Pongo*: upper molars lack lingual cingula and the third lower molar is smaller than M₂.

Schultz had defended the unity of the 'great' apes and vigorously opposed alternative theories, especially the grouping of man and African apes. Analysis of his enormous stock of morphological data on hominoids and Old World monkeys¹⁹⁻²¹

Table 1 Distribution among extant primates of some characters discussed in text

	<i>Homo</i>	<i>Pongo</i>	<i>Pan</i>	<i>Gorilla</i>	<i>Hylobates</i>	Cercopithecines	Colobines	Ceboids	Strepsirrhines
Nasal aperture	H-Tp	H-Tp	H-Tp	H-Tp	H-O	O	N, O	R-O	R-O
Subnasal plane	T	T	T	T	T	E	E	E	E
Subnasal plane stepped down to floor of nasal cavity	Variable	J	Variable	Variable	Present	Present	Present	Present	Variable
Orbits	N-Se	H	Variable	H	Se	B	B-Se	Se-H	Se-H
Interorbital distance	B	N	B	B	B	N	B	B	N-B
Infraorbital foramina: large	≤3	≤3	≤3	≤3	≤3	3+	≤3	≤3	≤3
Infraorbital foramina: distance from zygomaticomaxillary suture	Close	Far	Far	Far	Far	Close	Close	Variable	Variable
Zygomatic bone curved, with strong posterior slope	Present	Present	Present	Present	Present	Present	Present	Present	Variable
Zygomatic foramina	S	S-M	S-M	S	S-L	S-M	S-A	S-M	L-S-A
Zygomatic foramina	1-2	1-2+	1-2	1-2+	1-2	1-2	0-2	1-2	0-2+
Zygomatic foramina: relative to lower rim of orbits	Inf.	Sup.	Inf.	Inf.	Inf.	Inf.	Inf.	Inf.	Inf.
Incisive foramina	S-1	S-1	S-M-2	M-2	L-2	L-2	L-2	L-2	L-2
Greater palatine foramina	L, O	L, O	L, O	L, O	L, O	L, O	L, O	S, N	S, N
Upper incisors: size discrepancy	S	L	S	S	S	S	S	S	Variable
Enamel on molars	Thick	Thick	Thin	Thin	Thin	Thin*	Thin	Thin†	Thin
Cheek tooth height	Lo	Lo	M-Hi	Hi	M	M	M-Hi	M-Hi	M-Hi
Foramen lacerum	L	S-M	Absent	Absent	Absent	Absent	Absent	Absent	Absent

Distribution among extant primates of some characters discussed in text; these characters were originally studied by Andrews and Cronin¹⁰ but only among hominoids. The presence of relatively thick molar enamel in *Cebus* and *Cercopithecus* is a reflection of their autapomorphy within their own clades and is not indicative of this feature being primitive for *Homo* and *Pongo* or hominoids in general.

H, high; Tp, trapezoidal; O, oval-shaped; N, narrow; R, round; T, truncated; E, elongated; J, especially in juveniles; Se, subequal; B, broad; S, small; M, medium/moderate; L, large; Inf., inferior; Sup., superior.

* *Cercopithecus* has very thick enamel³.

† *Cebus apella* has very thick enamel³.

again sets apart man and orang-utans. For instance, among catarrhines, man and orang-utans are distinct in that they: (1) develop scapulae with vertical vertebral borders and small supra- but large infrapinnous fossae; (2) can grow the longest hair; (3) have the most widely separated, almost axillary, mammary glands; (4) have gestation periods of equal duration, which are also the longest among primates (now determined to be 270 days; R. D. Martin, personal communication). Schultz also pointed to the rare occurrence in *Pongo* of ischial callosities and to the fact that these structures, in contrast to those of *Pan* and *Gorilla*, are unkeratinized; ischial callosities are not found in humans.

Schultz's data on delayed epiphyseal development also reveal the broader pattern of nested groups of hominoids. All hominoids delay ossification of the distal humeral and proximal radial epiphyses until after birth. The large hominoids are distinguished further by the delay in ossification of the distal ulnar epiphysis and the distal metacarpal epiphyses of the hand, but *Homo* and *Pongo* are unique among extant hominoids in also delaying the ossification of the epiphyses of the proximal humerus and the distal radius. In addition, I have found that orang-utans, but not the African apes, develop a foramen lacerum basicranially, although it is smaller than in man, and Ward^{22,23} has demonstrated that *Pongo* shares with hominids upper canine rotation.

Recent studies of non-hard tissue features^{24,25} further distinguish man and orang-utans among hominoids. For instance, humans and orang-utans lack an oestrous cycle; copulation is not confined specifically (as in *Gorilla*) or generally (as in *Pan*) to the peak of the menstrual cycle; copulatory bouts are the longest among primates; and females have the highest levels of oestrial secretion during the menstrual cycle. Schultz also suspected that female orang-utans, like man, lack genital swelling, which has now been confirmed²⁶.

Various characters have been considered indicative of a man-African ape clade: (1) development of frontal sinuses; (2) prenatal fusion in the wrist of the os centrale and navicular bone; (3) broad incisive fossa; (4) restriction of the palatal ridges to in front of M¹; and (5) the occasional presence of earlobes. However, there are weaknesses in the association between some of these characteristics. (1) Schultz¹⁹ argued that, because the pygmy

chimpanzee rarely develops frontal sinuses, this was not a useful character. Cave and Haines²⁷ subsequently demonstrated that humans, *Gorilla* and the chimpanzee (*Pan troglodytes*) are, like man, distinguished in their neomorphic development of ethmoidal/frontal sinuses. (2) Gibbons are similar to man and African apes in that the two carpals fuse prenatally, but in orang-utans this occurs in middle-old age; these bones do not become fused in any other primate¹⁹. Fusion distinguishes the hominoids, but prenatal fusion does not single out humans and African apes. (3) Ward and Pilbeam²³ suggest that the presence of broad incisive fossae in the nasal cavity of African apes (especially *Pan*) and *Australopithecus* (especially *Australopithecus afarensis*) reflects the unity of hominids and African apes. These fossae, or canals, are extremely narrow in *Pongo* and *Homo sapiens*¹⁶. As the prevalent primate configuration is the persistence of relatively broad but patent fossae (palatine fenestrae)¹⁶, the broad fossae of the African apes and *Australopithecus* would be unique by comparison but, given the further reduction of this region in *Pongo* and *Homo*, primitive among the large hominoids. Together with sinus development, the disposition of the palatal ridges and the occasional development of earlobes remain only as potentially indicative of the affinities of humans and African apes. There appears to be only one character shared uniquely by humans and chimpanzees—the reduction of body hair, especially ventrally. Also, humans and gorillas are most similar in relative proportions, especially of the hands and feet.

Review of morphology reveals that when *Homo* shares a unique feature with another primate it is most frequently with the orang-utan. Given the systemic diversity of these features it is doubtful that they are components of a single functional complex and, as the gorilla is itself large but does not possess the suggested derived features, it is also reasonable to assume that the similarities between man and the orang-utan are not size-related. We could still decide that these features evolved in parallel in the orang-utan and *Homo*, but for this we must assume that the characters shared by man and the African apes reflect their closeness. Alternatively, the conclusion that the greater number of unique features shared by man and the orang-utan was inherited from a common ancestor not shared with any other extant hominoid seems stronger—in which case the characters suggestive of a possible human-African ape

Table 2 Distribution among extant hominoids of some uniquely shared morphologies

	<i>Homo</i>	<i>Pongo</i>	<i>Pan</i>	<i>Gorilla</i>	<i>Hylobates</i>
Distal humerus/proximal radius delayed ossification	X	X	X	X	X
Subnasal plane truncated	X	X	X	X	X
Ulnar-triquetral articulation variable	X	X	X	X	X
Infraorbital foramina removed from zygomaticomaxillary suture	X	X	X	X	X
Restriction enzyme <i>Ava</i> I at position 10, mtDNA	X	X	X	X	X
Nasal aperture trapezoidal	X	X	X	X	
Subnasal plane variably stepped-down	X	X	X	X	
Palatine fenestrae reduced in size	X	X	X	X	
Distal ulna/metacarpals delayed ossification	X	X	X	X	
Low-cusped cheekteeth	X	X			
Molar enamel thick	X	X			
Upper molars lacking cingula	X	X			
Lower M ₃ smaller than M ₂	X	X			
Single incisive foramen palatally	X	X			
Foramen lacerum	X	X			
Scapula short, vertical vertebral border, small superior fossa	X	X			
Proximal humerus/distal radius delayed ossification	X	X			
Ischial callosities unkeratinized, infrequent	X	X			
Longest hair	X	X			
Mammary glands widespread, c. axillary	X	X			
Longest gestation period: 270 days	X	X			
Female genitalia lack tumescence	X	X			
Copulation not confined to specific part of menstrual cycle	X	X			
Longest copulation bouts	X	X			
Highest oestriol levels during menstrual cycle	X	X			
Restriction enzyme (mtDNA position): <i>Hinc</i> II (21, 95), <i>Bam</i> HI (15), <i>Bgh</i> I (65), <i>Xho</i> I (10)	X	X			
Morphological correlates of knuckle-walking			X	X	

Distribution among extant hominoids of some uniquely shared morphologies. This is not exhaustive, but is representative of the degree to which *Homo* and *Pongo* are distinguished from the others. The extant hominoids are further united as a group by other characters, for example, dorsoventral compression of the thoracic cavity, dorsal emplacement of the scapula, elongation of the clavicle (see refs 17, 21). X, present.

relationship must be the parallelisms. Table 2 summarizes some of the morphologies shared more exclusively among extant hominoids.

Sivapithecus and the fossil record

The present controversy in hominoid systematics stems primarily from problems dealing with *Sivapithecus* which appears to have the jaws and teeth of a hominid but the face of an orang-utan. However, *Sivapithecus* shares many derived features with the orang-utan. These include narrow interorbital distance, flattened zygoma with relatively large foramina emplaced superior to the lower rim of the orbit, long slit-like incisive foramen, broadly continuous palate and premaxillary posterior pole, I¹ much larger than I², and P₃ metaconid region somewhat enlarged^{9,10,13,28}. These features would certainly seem to unite *Sivapithecus* and *Pongo*. Low-cusped cheekteeth and thick-enamelled molars have also been cited as uniting *Sivapithecus* and the orang-utan. But, since these dental traits are among those shared by *Homo* and *Pongo*, it would appear that they are present in *Sivapithecus* because it, too, belongs to the same clade: that is, *Sivapithecus* is related to the *Homo*-*Australopithecus* group by virtue of its relationship to *Pongo*. And this is why *Sivapithecus* shares derived features with both hominids and the orang-utan. *Gigantopithecus* also possesses the cheek-tooth specializations of the hominid-*Pongo*-*Sivapithecus* clade. Its specific affinities appear to lie with *Sivapithecus* and *Pongo*^{1,9}.

Non-morphological evidence

While many workers believe that molecular, biochemical and chromosomal data contribute to a resolution of phylogenetic problems, some believe that these procedures provide a greater specificity of phylogenetic detail (for example, refs 29-36), but morphologically-based phylogenies have been used to choose between statistically equivalent molecular phylogenies (for example, refs 29, 37). However, how does one choose the supporting morphologically-based hypothesis?

It is well known that contradictory results have long existed between certain molecular and biochemical 'schools' (for example, refs 38 versus 37, 39) and non-morphologists themselves have raised questions about, for example, determining homology⁴⁰, the structural level of the molecule being studied⁴¹, recognizing parallelism⁴², and even the use of protein data^{43,44}. In addition, there is mathematical objection to the validity of the maximum-parsimony method of assigning codons derived from sequence data (for example, see ref. 35) as well as to current determinations of immunological distance⁴⁵.

Many of the non-morphological studies that claim to demonstrate affinity between humans and one or both of the African apes lack data on all of the hominoids and/or also sample only a small number of the multitude of extant non-hominoid primate genera (for example, refs 31, 37, 38, 46-50). In such cases, the demonstration of similarity between humans and one or both of the African apes is phylogenetically meaningless if the goal is to argue that these hominoids are closely related. Even when the five hominoids are included in the sample, problems of understanding similarity persist. For example, man and the African apes have a 1.1% base pair mismatch in hybridized nuclear DNA; between humans and orang-utans it is 2.4% (ref. 51), but the significance of this difference is not clear, as nuclear DNA consists of ~10⁶-10⁷ base pairs. More importantly, however, one does not even know the nature of the difference. Is the orang-utan different because it has retained the ancestral hominoid nuclear DNA sequence whereas the DNA of humans and African apes has changed? Or is the orang-utan different because its nuclear DNA changed from the ancestral hominoid condition and humans and African apes have retained the primitive sequence? These are certainly equally valid interpretations and are applicable to all studies that attempt to deal with difference without knowing what the details of that difference are—for example, immunological studies that rely on relative degrees of antigenic reactivity (see refs 48, 52).

Although there is great potential in the analysis of mitochondrial (mt) DNA and nuclear DNA sequence data for elucidating phylogenetic relationships among taxa, because nucleotides are

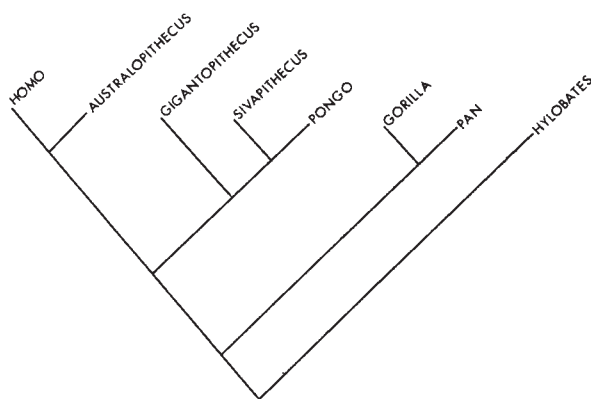


Fig. 1 Cladogram depicting the phylogenetic relationships among hominoids as discussed in the text. Characters that unite, respectively, the entire group, the large hominoids and the broader *Homo-Pongo* clade can be found in Table 2.

discrete units of molecular 'morphology', the body of data at present is quite limited⁵³. For example, mtDNA sequence data are known for the five extant hominoids, the mouse and the ox (see refs 49, 53). While it is reasonable to make outside (out-group) comparison in the determination of character states, a comparison between the five hominoid genera (the taxa of interest) and only two other mammals, especially in the absence of other primate as well as other mammalian taxa to serve as indicators of the polarity (primitiveness versus derivedness) of the character states determined for the hominoids, may not be helpful.

In general, statements of relatedness from sequence data are based on maximizing the (genetic) similarity of taxa. The assumption is that, the less time elapsed since the divergence of taxa, the more molecular similarity they will show. Uniting the chimpanzee with humans rather than with its fellow knuckle-walking African ape is indicated by '*Pan* sharing more base pair matches with *Homo* than with *Gorilla*'⁴⁹. But, as has been a major point of this review, overall similarity is not necessarily, nor should it be assumed to be, reflective of closeness of relatedness.

The pitfalls of such an assumption are exemplified by a study of α - and β -haemoglobins in which it was concluded that the identity in sequences of these molecules in the pygmy chimpanzee (*Pan paniscus*) and the chimpanzee (*Pan troglodytes*) as well as in *Homo* indicated not only the fact that the two chimpanzees are indeed sister-species but also that chimpanzees and humans are the most closely related among extant hominoids⁴⁹. But, given the lack of a sufficiently broad comparative sample, one cannot rule out the possibility that these three hominoids are identical in their haemoglobin sequences because of primitive retention and that the gibbon, gorilla and orang-utan are different, not because each is relatively more primitive, but because each has changed in its haemoglobin from the ancestral hominoid condition.

Study of restriction endonuclease cleavage site maps of mtDNA is another recent approach brought to bear on issues of phylogenetic relatedness (for example, ref. 30). Templeton's³⁶ analysis of Ferris *et al.*'s³⁰ data led to the conclusion that man-African ape and man-orang-utan theories of relatedness were both statistically viable; the choice between them was made by the use of other (that is, haemoglobin) data. This approach, as with most molecular and biochemical studies, is founded on presumptions of different rates of nucleotide substitution. However, as the determination of any rate of evolution is possible only after the phylogeny—the branching or successional sequence—has been agreed upon, it is doubtful that such studies could produce results greatly different from the earlier assumed phylogeny.

If we scrutinize the mtDNA data³⁰ in the same manner as morphological data, the following emerges: the extant hominoids are united by only one character, the presence of restriction enzyme *AvaI* (position 10); the gorilla and chimpanzee share *PstI* (position 32), *PvuII* (60, 70 and 95), *HpeI* (86), *HincII* (86) and *BamHII* (20); and the two species of *Pan* are distinguished by *XbaI* (16), *HindIII* (18 and 270), *PvuII* (35), *XbaI* (44 and 54), *HincII* (53), *FnuDII* (41 and 45), *BclI* (50 and 95), *AvaI* (6), *BamHI* (15), *BglI* (65) and *BstEII* (33). Humans and the African apes do share *BclI* (29) and *HincII* (86), but *Homo* and *Pongo* have in common the derived states *HincII* (21 and 95), *BamHI* (15), *BglI* (65) and *XhoI* (10). The relatedness of *Homo* and *Pongo* is suggested instead of that of *Homo* and the African apes. Some of these results are summarized in Table 2.

Conclusion

At present, it appears that non-morphological evidence does not necessarily support the theory that humans and African apes are closely related. However, some data do seem to point to the evolutionary closeness of humans and the orang-utan. This conclusion is also reached through the more traditional study of morphology. A consequence of the broader picture of hominoid evolution proposed here is that it is no longer necessary to explain why hominids, *Pongo* and *Sivapithecus* are so similar to each other in such seemingly unique features. A simple solution, and one that appears to be relatively well corroborated, is that hominids, *Pongo* and *Sivapithecus* inherited these features from a common ancestor not shared with any African ape (Fig. 1).

I thank Drs I. Tattersall and G. Musser (The American Museum of Natural History) and R. Thorington (Smithsonian Institution) for access to specimens in their charge and Drs E. Delson, T. Kaufman, L. Martin, A. Rogers, M. Salmon, A. Templeton and especially Drs P. Andrews, M. Goodman and R. D. Martin for comments critical to this version of the manuscript.

- Andrews, P. J. in *Recent Advances in Primatology* Vol. 3, 43–56 (Academic, New York, 1978).
- Kay, R. F. *Am. J. phys. Anthropol.* **43**, 195–215 (1975).
- Kay, R. F. *Am. J. phys. Anthropol.* **55**, 141–151 (1981).
- Molnar, S. & Gantt, D. G. *Am. J. phys. Anthropol.* **46**, 447–454 (1977).
- Simons, E. L. *Proc. natn. Acad. Sci. U.S.A.* **51**, 518–535 (1964).
- Simons, E. L. *J. hum. Evol.* **5**, 511–528 (1976).
- Simons, E. L. & Pilbeam, D. R. in *The Functional and Evolutionary Biology of Primates*, 36–62 (Aldine, Chicago, 1972).
- Smith, R. J. & Pilbeam, D. R. *Nature* **284**, 447–448 (1980).
- Andrews, P. J. *Nature* **295**, 185–186 (1982).
- Andrews, P. J. & Cronin, J. E. *Nature* **297**, 541–546 (1982).
- Andrews, P. J. & Tekkaya, I. *Palaeontology* **23**, 85–95 (1980).
- Pilbeam, D. R. *Nature* **295**, 232–234 (1982).
- Greenfield, L. O. *Am. J. phys. Anthropol.* **50**, 527–548 (1979).
- Kay, R. F. *Int. J. Primatol.* **3**, 113–173 (1982).
- Wolpoff, M. *Curr. Anthropol.* **23**, 501–522 (1982).
- Andrews, P. J. & Groves, C. *Gibbon Siamang* **4**, 167–218 (1975).
- Schwartz, J. H. *Primates* **24**, 231–240 (1983).
- Delson, E. & Andrews, P. J. in *Phylogeny of the Primates*, 405–446 (Plenum, New York, 1975).
- Schultz, A. H. *Q. Jl Biol.* **11**, 259–283, 425–455 (1936).
- Schultz, A. H. in *Classification and Human Evolution*, 85–115 (Aldine, Chicago, 1963).
- Schultz, A. H. in *Perspectives on Human Evolution* Vol. 1, 122–195 (Holt, Rinehart & Winston, New York, 1968).
- Ward, S. C. *Am. J. phys. Anthropol.* **60**, 268 (1983).
- Ward, S. C. & Pilbeam, D. in *New Interpretations of Ape and Human Ancestry*, 211–238 (Plenum, New York, 1983).
- Nadler, R. D. in *Reproductive Biology of the Great Apes*, 192–238 (Academic, New York, 1981).
- Collins, D. C., Graham, C. E. & Preedy, J. R. K. *Endocrinology* **96**, 93–101 (1975).
- Graham, C. E. in *Reproductive Biology of the Great Apes*, 1–43 (Academic, New York, 1981).
- Cave, A. J. E. & Haines, R. W. *J. Anat.* **74**, 493–523 (1940).
- Ward, S. C. & Kimbel, W. H. *Am. J. phys. Anthropol.* **61**, 157–171 (1983).
- Bruce, J. & Ayala, F. J. *Evolution* **33**, 1040–1056 (1979).
- Ferris, S. D., Wilson, A. C. & Brown, W. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2432–2436 (1981).
- Goodman, M. *Proc. Biophys. molec. Biol.* **37**, 105–164 (1981).
- Prager, E. M., Wilson, A. C., Lowenstein, J. M. & Sarich, V. M. *Science* **209**, 287–289 (1980).
- Lowenstein, J. M., Sarich, V. M. & Richardson, B. J. *Nature* **292**, 409–411 (1981).
- Prager, E. M. & Wilson, A. C. *Proc. natn. Acad. Sci. U.S.A.* **72**, 200–204 (1975).
- Selander, R. K. in *Perspectives on Evolution*, 32–59 (Sinauer, Sunderland, Massachusetts, 1982).
- Templeton, A. R. *Evolution* **37**, 221–244 (1983).

37. Baba, M., Weiss, M. L., Goodman, M. & Czelusniak, J. *Syst. Zool.* **31**, 156–165 (1982).
38. Sarich, V. M. & Cronin, J. E. in *Molecular Anthropology*, 141–170 (Plenum, New York, 1976).
39. Dene, H. T., Goodman, M. & Prychodko, W. in *Molecular Anthropology*, 171–196 (Plenum, New York, 1976).
40. Williams, C. A. Jr *Ann. N.Y. Acad. Sci.* **102**, 249–252 (1962).
41. Romero-Herrera, A. E., Lehmann, H., Joysey, K. A. & Friday, A. E. *Phil. Trans. R. Soc. B* **283**, 61–183 (1978).
42. Doolittle, R. F. *Science* **214**, 149–159 (1981).
43. Adams, J. & Rothman, E. D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3560–3564 (1982).
44. Diamond, J. M. *Nature* **305**, 17–18 (1983).
45. Farris, J. S. in *Advances in Cladistics*, 3–23 (New York Botanical Gardens, 1981).
46. Goodman, M. in *Phylogeny of the Primates*, 219–248 (Plenum, New York, 1975).
47. Stanyon, R. & Chiarelli, B. *J. hum. Evol.* **11**, 493–504 (1982).
48. Yunis, J. J. & Prakash, O. *Science* **215**, 1525–1530 (1982).
49. Goodman, M., Braunitzer, G., Stang, A. & Schrank, B. *Nature* **303**, 546–548 (1983).
50. Mai, L. L. in *New Interpretations of Ape and Human Ancestry*, 87–114 (Plenum, New York, 1983).
51. Benveniste, R. E. & Todaro, G. J. *Nature* **261**, 101–108 (1976).
52. Bonney, R. C. & Kingsley, S. *Int. J. Primatol.* **3**, 431–444 (1983).
53. Goodman, M., Olson, C. B., Beeber, J. E. & Czelusniak, J. *Acta zool. fenn.* **169**, 19–35 (1982).

ARTICLES

The olivine → spinel transformation and the rheology of subducting lithosphere

David C. Rubie*

Departement für Erdwissenschaften, ETH-Zentrum, 8092 Zürich, Switzerland

During subduction, relatively hot lithospheric slabs are expected to deform in a superplastic flow regime below a depth of ~300 km if olivine reacts to fine-grained spinel. This process reduces the flow strength of such slabs below the phase boundary so that they become aseismic at about this depth. In colder slabs temperatures are too low for superplastic deformation, and a high flow strength and seismic activity are maintained to a depth of 600–700 km.

THE scale of mantle convection and the fate of subducting lithosphere at the 670 km discontinuity are major factors controlling the long-term geochemical evolution of the mantle and lithosphere. Although Ringwood has argued that most of a subducted slab is unable to penetrate to below a depth of 670 km¹, other theoretical studies have been less conclusive². This uncertainty largely arises because the physical and chemical characteristics of the 670 km discontinuity are poorly known parameters in models of the mantle^{2–4}. However, another parameter, which has received less attention, is the rheological state of the slab and its effect on slab dynamics. Here I propose that a significant change in rheology occurs in some slabs at a depth of ~300 km as a result of the transformation of olivine → spinel (β and/or γ polymorphs).

Subduction zone seismicity

Direct information on the mechanical properties of subducting lithosphere originates from seismic data. On the basis of spatial patterns of intermediate and deep earthquake activity^{5,6}, most subduction zones fall into the three groups shown in Fig. 1. In group 1 subduction zones, earthquake activity is absent below a depth of 250–300 km. In group 2, earthquakes occur to a depth of ~600 km but with a distinct gap in seismic activity between ~300 km and ~500 km. In group 3, seismic activity is continuous to a depth of ≥500–600 km. There is also some evidence that group 3 slabs may be further subdivided into those that penetrate the 670 km discontinuity (for example, Kuriles subduction zone⁷), and those which do not (for example, Tonga subduction zone⁸).

Previously, attempts have been made to relate some of these seismicity characteristics to the thermal structure of the slab^{9,10}. Two explanations have been proposed for the aseismic gap in slabs of group 2: (1) It is a depth interval over which there is a switch from down-dip extension to down-dip shortening of the slab (Fig. 1), that is, a zone of low differential stress^{5,6}. (2) The lower part of the lithospheric slab has become detached, thus creating a physical gap in the slab¹¹. A third explanation, which is still compatible with a change-over of the down-dip component of stress, is presented below.

Oceanic lithosphere is thought to have a layered structure consisting of basaltic crust, harzburgite, lherzolite and pyrolite¹. Except in the thin basaltic layer, olivine (Fo₉₀) forms between ~60% and ~80% of these lithologies. It is expected, therefore, that the mechanical properties of olivine (α) and of its high-pressure spinel polymorphs (β and γ) must exert an important influence on the rheology of a subducting slab to a depth of 670 km. The depth at which olivine transforms to spinel depends on the thermal structure of a slab (Fig. 2) and therefore on its age and rate of subduction⁹. Estimates of this depth for the cool interior of a slab vary from ~280 km¹² to ~350 km¹. It is notable that these estimates closely correspond to the depths of the lower limit of seismic activity in slabs of group 1 and to the upper limit of the gap in seismic activity in slabs of group 2 (Fig. 1). I will therefore discuss the hypothesis that slabs can become either temporarily or permanently aseismic as a consequence of the olivine → spinel transformation.

Rheology of subducting lithosphere

If the flow strength of a slab is high, deformation in response to the accumulation of large stresses is likely to be localized on fault zones, perhaps comparable to those produced experimentally in dunite^{13,14}. Catastrophic failure on such fault zones may be responsible for most, if not all, intermediate and deep earthquakes. If the olivine → spinel transformation results in either no change or an increase in flow strength, then failure on fault zones, and associated seismic activity, should continue at depths below the phase boundary. If, however, olivine reacts to an aggregate of fine-grained spinel, deformation by grain size-sensitive mechanisms of structural superplasticity¹⁵ may occur below the phase boundary, with a consequent decrease in flow strength. In this case the slab would become aseismic if the decrease in flow strength was sufficient to preclude catastrophic failure. Similarly, the absence of earthquakes below 700 km could be due to superplastic deformation of fine-grained products resulting from spinel breakdown reactions¹⁶. There is also petrographic evidence that the flow strength of some crustal rocks can undergo a considerable reduction during continental collision as a result of the reaction of albite to fine-grained (superplastic) jadeite + quartz¹⁷.

Mechanisms of structural superplasticity are thought to be dominated by diffusion-accommodated grain boundary sliding

* Present address: Mineralogisch-Petrographisches Institut, Universität Bern, Baltzerstrasse 1, CH-3012 Bern, Switzerland.

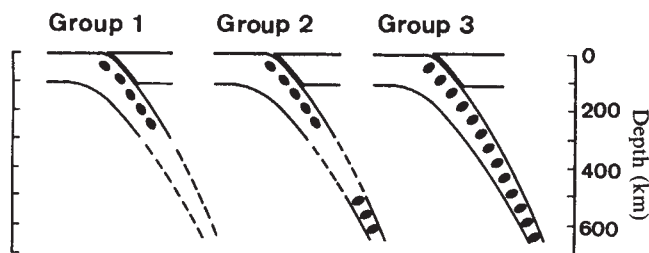


Fig. 1 Three groups of subduction zones classified on the basis of seismicity. Seismically active parts of a slab are indicated by solid lines and by schematic strain ellipses. The latter indicate parts of a slab which are undergoing either down-dip extension or down-dip shortening. Dashed lines indicate regions where a slab is either aseismic or absent. (Modified from ref. 5, Figs 2 and 3.)

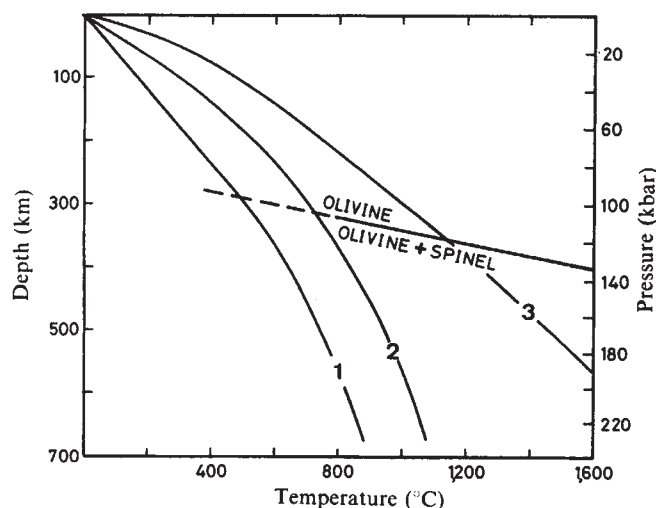


Fig. 2 Temperature-depth diagram showing calculated minimum-temperature profiles in subducting slabs. Curve 1 is from ref. 12, curve 2 from ref. 28, and curve 3 from ref. 40. The different profiles result from the use of different parameters and boundary conditions, but a similar variation is obtained by varying the rate of subduction⁴¹. The olivine-breakdown curve is for a composition of Fo₈₈ (ref. 42). T^* varies from 500 °C to 1,100 °C for the profiles shown.

in many cases¹⁵. Provided the grain size of reaction products remains small (that is, little or no grain growth), a low flow strength and the development of high strains can be maintained in a material deforming by such a mechanism after a phase transformation has reached completion¹⁷. This is in marked contrast to the phenomenon of transformation plasticity/superplasticity which can only influence rheology while a phase transformation is in progress^{15,18}. Depending on the kinetics of the olivine → spinel transformation and the pressure-temperature conditions in a slab, transformation plasticity may only affect the rheology of subducting lithosphere over a very restricted depth interval. The possible role of transformation plasticity is not considered here, and henceforth the term 'superplasticity' is used as being synonymous with 'structural superplasticity'.

Fine-grained aggregates of reaction products have resulted during some experimental studies of high-pressure polymorphic phase transformations^{19,20}. The experiments of Vaughan and Coe on the analogues Mg₂GeO₄ olivine and Mg₂GeO₄ spinel are of relevance to the mantle²⁰. A hot-pressed Mg₂GeO₄ olivine aggregate of grain size 30–200 μm was reacted to Mg₂GeO₄ spinel with a grain size of ~3 μm diameter. From the results of their deformation experiments, Vaughan and Coe concluded, after careful discussion, that olivine probably deformed by dislocation creep, whereas the fine-grained spinel deformed in a superplastic flow regime²⁰.

Estimates of the change in flow strength across the olivine → spinel phase boundary in a subducting slab can be made using empirical flow data for Mg₂SiO₄ olivine²¹ and Mg₂GeO₄

Table 1 Estimates of subduction zone parameters

	Age (Myr)*	Rate (cm yr ⁻¹)	10 ³ /[(age × rate) [10 ³ /(Myr cm yr ⁻¹)]
Group 1			
Ryukys	135	4.0†	1.85
Hokkaido	100	5.6†	1.79
Sumatra	38–135	6.5‡	4.05–1.14
S. Mexico/Central America	40	8.0–8.5†‡§	3.13–2.94
Aleutians	60	5.2–8.9‡§	3.21–1.87
Ecuador/N. Peru	45	8.2‡	2.70
Altiplano	45	9.1‡	2.44
			Average = 2.51
Group 2			
Java	100	6.7‡	1.49
Kamchatka	100	6.3‡	1.59
New Hebrides	50	7.0–9.8‡§	2.86–2.04
N. Chile	45	9.0–9.5†‡	2.47–2.34
New Zealand	100	4.6–6.4‡	2.17–1.56
			Average = 2.07
Group 3			
Mindanao	135	7.9‡	0.94
Kermadec	100	6.4–7.7†‡§	1.56–1.30
Marianas	135	5.6–9.1†‡	1.32–0.81
Izu-Bonin	100	7.4–9.8†‡§	1.35–1.02
Kuriles	135	7.8–10.7‡§	0.95–0.69
Tonga	100	6.1–9.5†‡	1.64–1.05
			Average = 1.15

* Ref. 36; † ref. 37; ‡ ref. 38; § ref. 39; || ref. 9.

spinel²⁰. It is necessary to assume that Mg₂GeO₄ spinel is quantitatively a reasonable analogue for Mg₂SiO₄ spinel; the similarity between dislocation creep data for Mg₂SiO₄ olivine and Mg₂GeO₄ olivine lends some support to this assumption^{20–22}. It must be emphasized that the uncertainties in these estimates are large due to the need for making large extrapolations in strain rate and due to uncertainties in the parameters discussed below.

Ductile deformation of olivine can occur in a high temperature/low stress regime characterized by dislocation creep, or in a low temperature/high stress regime in which glide dominates (low-temperature plasticity)^{21,23}. Empirical flow laws for these regimes have been published²¹. To estimate the flow stress of a subducting slab in the vicinity of the olivine → spinel phase boundary, it is necessary to extrapolate these flow laws to high pressures. For olivine, this was done by following the procedure of Ashby and Verrall for the low-temperature plasticity regime²⁴, and by using an activation volume of 11 cm³ mol⁻¹ for dislocation creep²⁵. In a given set of conditions, the flow stress can be estimated from the law which predicts the lowest strength. In the conditions of high pressure and relatively moderate temperature immediately above the olivine → spinel phase boundary in the interior of a slab ($T = 600–900$ °C, $P \sim 120$ kbar), a considerably lower flow stress is predicted by the low-temperature plasticity flow law than by the dislocation creep flow law. The flow stress (σ_{OI}) of olivine has therefore been calculated using Goetze's law for the low-temperature plasticity regime²¹:

$$\dot{\epsilon} = 5.7 \times 10^{11} \exp \left[\frac{-128 \text{ kcal mol}^{-1}}{RT} \left(1 - \frac{\sigma_{OI}}{85} \right)^2 \right] \quad (1)$$

where $\dot{\epsilon}$ is the strain rate, σ_{OI} is flow stress (kbar), R is the gas constant and T is the temperature (K). Equation (1) can be rewritten to incorporate pressure-dependent terms²⁴:

$$\dot{\epsilon} = 5.7 \times 10^{11} \exp \left[\frac{-128 \mu b^3 \text{ kcal mol}^{-1}}{\mu_0 b_0^3 RT} \left(1 - \frac{\sigma_{OI} \mu_0}{85 \mu} \right)^2 \right] \quad (2)$$

where μ and b are the shear modulus and Burger's vector, respectively, at pressure P , and μ_0 and b_0 are the same para-

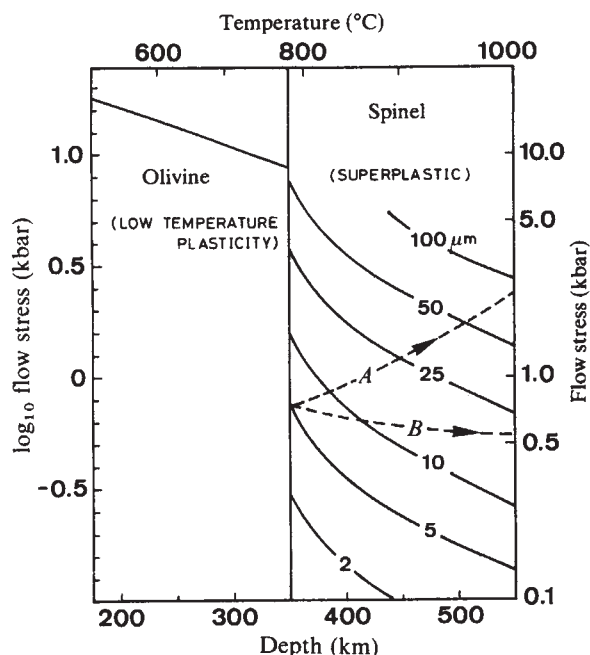


Fig. 3 The variation of flow stress with depth in the interior of subducting lithosphere for Mg_2SiO_4 olivine (<350 km depth) and Mg_2GeO_4 spinel (>350 km depth) for the temperature–depth profile calculated by Hsui and Toksöz (ref. 28, Fig. 9). The flow stress was calculated using equation (2) for low-temperature plasticity in olivine, and equation (4) for superplastic flow of spinel, using a strain rate of 10^{-14} s^{-1} . The spinel field is contoured for grain size (μm). Two schematic grain size/depth paths (A and B) are shown, assuming grain growth of initially fine-grained spinel. The actual grain size/depth path depends on the kinetics of grain growth, which are a function of temperature and pressure.

meters at 1 atm (P_0). μ and b are estimated from²⁴:

$$\mu = \mu_0 + \frac{d\mu}{dP}(P - P_0)$$

and

$$b = b_0 \exp \left(-\frac{(P - P_0)}{3K_0} \right)$$

where K_0 is the bulk modulus at 300 K and 1 atm. Values of b_0 , μ_0 , K_0 and $d\mu/dP$ were taken from Ashby and Verrall²⁴. As equations (1) and (2) are derived from yield stress data, they do not include the effects of work hardening, thus they predict only a minimum value of σ_{OI} .

The flow stress, σ_{Sp} , of fine-grained spinel has been estimated from the superplastic flow law of Vaughan and Coe for Mg_2GeO_4 spinel of grain size diameter $\sim 3 \mu\text{m}$, at a pressure of $\sim 18 \text{ kbar}$ ²⁰.

$$\dot{\epsilon} = 2.6 \times 10^4 \sigma_{\text{Sp}}^2 \exp \left(\frac{-73 \text{ kcal mol}^{-1}}{RT} \right) \quad (3)$$

where σ_{Sp} is flow stress in kbar. It is assumed that this law obeys the relationship $\dot{\epsilon} \propto 1/d^2$ where d is grain diameter¹⁵. The activation volume, V^* , for lattice diffusion in spinel is estimated to be $8 \text{ cm}^3 \text{ mol}^{-1}$ from data for olivine²⁵ and from a theoretical estimate of the change in activation volume across a phase boundary²⁶. It is further assumed that superplastic deformation of spinel is controlled by grain boundary diffusion. No experimental data for the activation volume (V_b^*) for grain boundary diffusion are available, but theoretical estimates of V_b^* for metals range from $0.4 V^*$ to $0.65 V^*$ (ref. 27). An estimate of V_b^* for spinel of $2 V^*/3$ ($= 127 \text{ cal kbar}^{-1} \text{ mol}^{-1}$) is used here. Incorporating these estimates into equation (2) gives:

$$\dot{\epsilon} = \frac{2.34 \times 10^5}{d^2} \sigma_{\text{Sp}}^2 \exp \left[\frac{-(70.7 + 0.127P) \text{ kcal mol}^{-1}}{RT} \right] \quad (4)$$

where P is the pressure (kbar) and d is in μm .

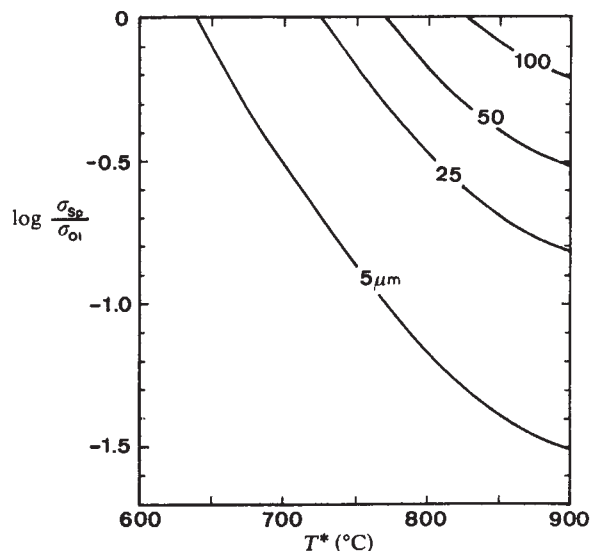


Fig. 4 The ratio of the calculated flow stresses of fine-grained Mg_2GeO_4 spinel (σ_{Sp}) and Mg_2SiO_4 olivine (σ_{OI}) at the olivine→spinel phase boundary in a subducting slab, plotted against temperature (T^*) at the phase boundary. The contours are for spinel grain diameter (μm). The magnitude of stress reduction across the phase boundary depends on the spinel grain size and on T^* .

The calculated flow stresses of Mg_2SiO_4 olivine and Mg_2GeO_4 spinel in the cool interior of a subducting slab are plotted against depth in Fig. 3 using a temperature–depth profile calculated by Hsui and Toksöz (Fig. 9 of ref. 28). The spinel field is contoured for grain size. Along this particular temperature–depth profile, the olivine→spinel phase boundary is expected to occur at $\sim 800^\circ\text{C}$ and at a depth of $\sim 350 \text{ km}$, assuming that some overstepping of equilibrium boundaries is required for nucleation (Fig. 2). Although Fig. 3 cannot be relied on quantitatively, due to large uncertainties in the assumptions listed above, it does suggest two conclusions: (1) flow stresses above the olivine→spinel phase boundary calculated from equation (2) are very high. It can be argued, therefore, that deformation of this part of the slab is not likely to develop homogeneously, but is likely to occur in localized fault zones, and probably in response to lower stresses than those shown in Fig. 3. Seismic activity must result if the development of such fault zones is catastrophic. (2) A significant (order-of-magnitude) decrease in flow stress across the phase boundary is possible if olivine reacts to spinel with $d \sim 5 \mu\text{m}$. This could be sufficient to preclude catastrophic failure below the phase boundary, it would also result in the elimination of high stresses in the slab below the phase boundary, and perhaps in the development of high strains.

Variations in rheology

The temperature, T^* , at which the olivine→spinel phase boundary is crossed in the interior of a subducting slab must vary considerably between different slabs (Fig. 2). In a rapidly subducting slab consisting of old lithosphere, T^* in the slab interior may be as low as 600°C (ref. 12). In the interior of a slowly subducting slab consisting of young lithosphere, T^* could be much higher (for example, 900°C). The thermal structure of a slab, and therefore T^* , can be related approximately to the rate of subduction and to the age of the lithosphere by the relationship⁹ $T^* \propto 1/(\text{rate} \times \text{age})$. Values of $1/(\text{rate} \times \text{age})$ for subduction zones provisionally classified into the three groups of Fig. 1, are listed in Table 1. The data suggest that T^* is highest in subduction zones of group 1 and lowest in those of group 3.

The estimated change in flow strength across the phase boundary is strongly dependent on T^* . The ratio of the calculated flow stresses, $\sigma_{\text{Sp}}/\sigma_{\text{OI}}$, of Mg_2GeO_4 spinel [equation (4)] and Mg_2SiO_4 olivine [equation (2)] is plotted against T^* in Fig. 4 (with $\dot{\epsilon} = 10^{-14} \text{ s}^{-1}$). For a spinel grain size of $5 \mu\text{m}$, Fig. 4 predicts no reduction in flow strength across the phase boundary

in relatively cold slabs (group 3) when $T^* \leq 640^\circ\text{C}$. In this case, seismic activity is correctly predicted to continue to depths exceeding 300–400 km. At higher values of T^* (groups 1 and 2), for example, 850°C , the calculated flow stress is reduced by more than an order of magnitude (to ~ 200 bar) if $d = 5\ \mu\text{m}$; this may explain why slabs of groups 1 and 2 become aseismic at depths of 250–300 km (Fig. 1). The magnitude of the reduction in flow stress also depends critically on the grain size of spinel (Figs 3, 4). For a particular grain size there is a critical value of T^* below which there is no reduction in flow strength.

Fine-grained single-phase aggregates are microstructurally unstable at high temperatures and are affected by the process of grain growth²⁹. Along a temperature–depth profile such as the one used to construct Fig. 3, unless the kinetics of grain growth of spinel are very slow (for example, Fig. 3, path B), grain growth of this material would result in a progressive increase in flow strength of subducting lithosphere with depth below the olivine $\rightarrow$ spinel phase boundary (for example Fig. 3, path A). At some depth, which depends on the kinetics of grain growth, a slab could again become seismically active due to an increased flow strength. Thus, reaction of olivine to fine-grained spinel, followed by grain growth or another hardening process, may explain the aseismic gap between ~ 300 km and ~ 500 km in slabs of group 2. When experimental data on the kinetics of grain growth of spinel at high pressure become available, it should be possible to test this hypothesis.

Slabs of group 1 have the highest values of T^* and remain aseismic at all depths below the olivine $\rightarrow$ spinel phase boundary. Earthquakes would be absent below a depth of 300 km if a low flow strength is maintained with increasing depth of subduction, even though rates of grain growth must be faster than in group 2 slabs. Due to the comparatively high temperatures of group 1 slabs, there may be a transition from a superplastic flow regime to a high temperature/low stress dislocation creep regime at some depth below the phase boundary. Furthermore, if the flow strength below the phase boundary is low, the thickness of the slab could decrease considerably due to the development of high strains. This would speed up thermal equilibration with the surrounding mantle, thus making a transition to a dislocation creep regime even more likely.

Related problems

Several basic questions arise from this discussion for which there seem to be no answers at present.

(1) Which parameters control the grain size of the products of a solid–solid transformation when the reaction reaches completion? It is possible that the rate of change of pressure and/or temperature is important, with rapid rates of change favouring fine-grained products¹⁷. As shown elsewhere¹⁷, subduction involves much faster rates of change of (P , T) than other tectonic processes. However, the grain size must also be related to the nucleation rate and the reaction mechanism. Because of the unresolved controversy as to whether the olivine $\rightarrow$ spinel trans-

formation proceeds by a nucleation and growth or a martensitic (diffusionless) mechanism^{30–34}, it is impossible at present to predict whether rates of diffusion of oxygen or silicon have any effect on grain size. The volume change associated with the transformation may have some control over grain size. The volume change resulting from the growth of a nucleus must be accommodated by plastic deformation of the reactant phase—this involves the formation and movement of defects or microfractures within the reactant³² which may act as sites for further nucleation. If the volume change is large (for example, 10%) a high defect density³², and therefore a high nucleation rate, may be maintained by this mechanism until the transformation is complete, and consequently the reaction products would be fine-grained.

(2) What are the kinetics of grain growth of a single-phase spinel aggregate at high temperatures (600–1,000 $^\circ\text{C}$) and very high pressures (100–160 kbar)? Presently available data are contradictory. Hamaya and Akimoto reported extremely rapid grain growth of Ni_2SiO_4 spinel at 1,000 $^\circ\text{C}$ and 37 kbar, with the formation of crystals of 300 μm diameter in 20 min³⁴. In contrast, during the experiments of Vaughan and Coe on Mg_2GeO_4 spinel, a small (3 μm diameter) grain size was apparently relatively stable, in conditions of both hydrostatic and non-hydrostatic stress, at 900–1,100 $^\circ\text{C}$ and 18 kbar for over 2 days²⁰. Data are available for the kinetics of grain growth of quartz aggregates up to 15 kbar³⁵. However, as the experiments were done in the presence of excess water, which greatly enhances growth rates, it is doubtful whether the data can be extrapolated to mantle conditions, even if quartz were an acceptable analogue for spinel. Most other experimental studies of grain growth kinetics, of metals and ceramics, have been done at atmospheric pressure. Assuming that the activation volume for grain growth is positive, rates should be significantly slower at very high pressures.

Thus, the rheologies of different slabs (for example, groups 1 and 3) are expected to be quite different over the 300–670 km depth interval. This may mean that the dynamics of these slabs will also be different. It is possible that cold rigid slabs (for example, group 3) penetrate the 670 km discontinuity, whereas hot ductile slabs (for example, group 1) do not. There is evidence that the slab penetrates the 670 km discontinuity in the Kuriles subduction zone⁷, whereas in the Tonga subduction zone there is apparently no such penetration⁸—both these subduction zones are in group 3. However, the data of Table 1 suggest that temperatures should be lower in the Kuriles slab than in the Tonga slab. This may be an example of how the thermal structure of a slab controls rheology, which in turn controls the ability of the slab to penetrate the 670 km discontinuity.

I thank U. Christensen, R. J. O'Connell, B. K. Smith, A. B. Thompson, J. Tullis, D. A. Yuen and R. A. Yund for valuable comments, and D. Giardini and J. H. Woodhouse for providing a preprint. The work was supported by ETH Forschungskredit (no. 5.501.330.752/4).

Received 22 September 1983; accepted 12 January 1984.

- Ringwood, A. E. *J. Geol.* **90**, 611–643 (1982).
- Christensen, U. & Yuen, D. A. *Trans. Am. geophys. Un.* **64**, 311 (1983).
- Jeanloz, R. & Thompson, A. B. *Rev. Geophys. Space Phys.* **21**, 51–74 (1983).
- Lees, A. C., Bukowski, M. S. T. & Jeanloz, R. *J. geophys. Res.* **88**, 8145–8159 (1983).
- Isacks, B. & Molnar, P. *Rev. Geophys. Space Phys.* **9**, 103–174 (1971).
- Vassiliou, M. S. thesis, California Inst. Technol., Pasadena (1983).
- Jordan, T. H. & Creager, K. C. *Trans. Am. geophys. Un.* **63**, 1254–1255 (1982).
- Giardini, D. & Woodhouse, J. H. *Nature* **307**, 505–509 (1984).
- Molnar, P., Freedman, D. & Shih, J. S. F. *Geophys. J. R. astr. Soc.* **56**, 41–54 (1979).
- Vlaar, N. J. & Wortel, M. J. R. *Tectonophysics* **32**, 331–351 (1976).
- Barazangi, M., Isacks, B. L., Oliver, J., Dubois, J. & Pascal, G. *Nature* **242**, 98–101 (1973).
- Schubert, G., Yuen, D. A. & Turcotte, D. L. *Geophys. J. R. astr. Soc.* **42**, 705–735 (1975).
- Blacic, J. D. *Am. geophys. Un. Monogr.* **16**, 109–115 (1972).
- Post, R. L. Jr *Tectonophysics* **42**, 75–110 (1977).
- Edington, J. W., Melton, K. N. & Cutler, C. P. *Prog. Mater. Sci.* **21**, 61–158 (1976).
- O'Connell, R. J. *Tectonophysics* **38**, 119–136 (1977).
- Rubie, D. C. *Tectonophysics* **96**, 331–352 (1983).
- Poirier, J. P. *J. geophys. Res.* **87**, 6791–6797 (1982).
- Wenk, H. R., Venkatasubramanian, C. S. & Baker, D. W. *Contr. Miner. Petrol.* **38**, 81–114 (1973).
- Vaughan, P. J. & Coe, R. S. *J. geophys. Res.* **86**, 389–404 (1981).
- Goetze, C. *Phil. Trans. R. Soc. A* **288**, 99–119 (1978).

- Vaughan, P. J. & Coe, R. S. *Tectonophysics* **46**, 187–196 (1978).
- Stocker, R. L. & Ashby, M. F. *Rev. Geophys. Space Phys.* **11**, 391–426 (1973).
- Ashby, M. F. & Verrall, R. A. *Phil. Trans. R. Soc. A* **288**, 59–95 (1977).
- Kohlstedt, D. L., Nichols, H. P. K. & Hornack, P. J. *geophys. Res.* **82**, 3747–3761 (1977).
- Sammis, C. G., Smith, J. C., Schubert, G. & Yuen, D. A. *J. geophys. Res.* **86**, 10707–10718 (1981).
- Brown, A. M. & Ashby, M. F. *Acta metall.* **28**, 1085–1101 (1980).
- Hsui, A. T. & Toksöz, M. N. *Tectonophysics* **60**, 43–60 (1979).
- Martin, J. W. & Doherty, R. D. *Stability of Microstructure in Metallic Systems* (Cambridge University Press, 1976).
- Sung, C. M. & Burns, R. G. *Phys. Chem. Miner.* **2**, 177–197 (1978).
- Vaughan, P. J., Green, H. W. II & Coe, R. S. *Nature* **298**, 357–358 (1982).
- Boland, N. J. & Lieberman, R. C. *Trans. Am. geophys. Un.* **63**, 431 (1982).
- Poirier, J. P. *Am. geophys. Un. Monogr.* 113–117 (1981).
- Hamaya, N. & Akimoto, S. *Phys. Earth planet. Interiors* **29**, 6–11 (1982).
- Tullis, J. & Yund, R. A. *J. Geol.* **90**, 301–318 (1982).
- Pitman, W. C. III, Larson, R. L. & Herron, E. M. *The Age of the Ocean Basins and Magnetic Lineations of the Oceans* (Am. geol. Soc., Boulder, Colorado, 1974).
- Yokokura, T. *Tectonophysics* **77**, 63–77 (1981).
- Uyeda, S. & Kanamori, H. *J. geophys. Res.* **84**, 1049–1061 (1979).
- Furlong, K. P., Chapman, D. S. & Alfeld, P. W. *J. geophys. Res.* **87**, 1786–1802 (1982).
- Toksöz, M. N., Sleep, N. H. & Smith, A. T. *Geophys. J. R. astr. Soc.* **35**, 285–310 (1973).
- Minear, J. W. & Toksöz, M. N. *J. geophys. Res.* **75**, 1397–1419 (1970).
- Akaogi, M. & Akimoto, S. *Phys. Earth planet. Interiors* **19**, 31–51 (1979).

The locus of sequence-directed and protein-induced DNA bending

Hen-Ming Wu & Donald M. Crothers

Chemistry Department, Yale University, New Haven, Connecticut 06511, USA

The bending locus of trypanosome kinetoplast DNA, identified by gel electrophoresis, has tracts of a simple repeat sequence (CA₅₋₆ T) symmetrically distributed about it, with a repeat interval of 10 base pairs. The analogous bending induced when catabolite gene activating protein binds to its recognition sequence near the promoter of the Escherichia coli lac operon is centred on a site about 5-7 base pairs away from the centre of the protein binding site.

BENDING is a known conformational change of the DNA double helix. DNA may bend as a result of interaction with proteins, for example in nucleosomes¹, gyasomes², intersomes³ and possibly the *Cro* protein-DNA complex⁴. Smaller ligands such as distamycin may also bend DNA⁵. DNA can assume a bent conformation by itself, as in supercoiled DNA and trypanosome kinetoplast DNA⁶. Proposed molecular models for bending include smooth curvature along the helix, and localized bends such as at kinks¹ or junction bends⁷⁻⁹.

Bending of DNA in response to regulatory protein binding is a largely unexplored phenomenon in the mechanism of gene control. A widely studied model system is transcription of the *Escherichia coli lac* operon, which is stimulated by interaction of the cyclic AMP receptor and gene activating protein (CAP or CRP) with a DNA binding site near the promoter^{10,11}. The molecular mechanism of stimulation of this classical gene regulatory system is still not known. Two general models have been proposed: the first suggests that direct interaction between CAP and RNA polymerase results in stronger binding of polymerase to the promoter¹²⁻¹⁶; in the second model a DNA structural transition, resulting from the binding of CAP, alters the structure or stability of the double helix in the adjacent RNA polymerase binding site¹⁷⁻²². The large variation in the distance between the CAP and RNA polymerase binding sites in some other catabolite-sensitive operons such as *gal*²³, *ara*²⁴ and *deo*²⁵ argues against a unique direct CAP-RNA polymerase interaction. The finding of not more than a half helical turn unwound in the *lac* promoter region²⁶ by CAP binding and the stabilization of the same region against DNA melting²⁷ upon binding with CAP, argue against a drastic CAP-induced DNA structural alteration in this region.

Study of possible bending of DNA fragments by proteins has suffered from the lack of experimental methods to detect the phenomenon. Recently, DNA fragments isolated from the kinetoplast body of trypanosome parasites have provided a model system for exploring the properties of bent DNA⁶. Especially striking is the anomalously slow gel electrophoretic mobility exhibited by kinetoplast DNA fragments, presumably because the molecules of bent conformation have increased difficulty passing through gel pores which are not much larger than the DNA helix diameter²⁸. In addition, the bent fragment exhibited an accelerated rotational relaxation time, as expected for DNA molecules rendered more compact by bending.

The studies reported here allow us to identify the DNA sequence at the bending locus of kinetoplast (K) DNA. This is achieved by comparing the gel mobility of a set of circularly permuted K-DNA fragments having a variable position of the bending site relative to the fragment ends. A periodic distribution of CA₅₋₆T tracts is found around the bend centre, with repeat equal to that of the DNA helix. Analogous studies demonstrate DNA bending in the CAP-DNA complex, and place the bend centre about 5-7 base pairs towards the promoter from the centre of the protein binding site. Models for DNA bending and promoter activation are presented.

Conformation of circular DNA fragments

Figure 1a illustrates the basic concept of our experiment. Separate samples of a cloned tandem dimer of a DNA fragment containing a bend are cleaved with enzymes which cut only once in the sequence. The result is a set of fragments whose conformations differ because of variation in the position of the bend relative to the molecular ends. When the bend is near the middle of the molecule, as in sample (1), the strongly bent overall shape should encounter more difficulty in snaking through the pores of a polyacrylamide gel than is experienced by the more linear fragment (2). This qualitative idea is supported by theories of gel electrophoresis^{29,30}, which predict a dependence of mobility on mean square end-to-end distance, yielding a lower predicted mobility for fragment (1).

Figure 1b illustrates of the dependence of gel mobility on position of the bend. A circular plasmid DNA containing a 423 base pair (bp) kinetoplast DNA insert was cleaved once, either close to or far from the K region. The results show that even in a large DNA molecule, significantly lower electrophoretic mobility is observed as the bending locus is moved nearer the centre of the molecule.

It is noteworthy that even the control plasmid vector pJW37 (a derivative of pBR322) showed small but significantly different mobilities for the different circularly permuted linear plasmids obtained by its cleavage. This indicates that some part of this plasmid may also exhibit bending, although less pronounced than shown by K-DNA. Sequence-directed DNA bending is probably a general phenomenon for which K-DNA provides a dramatic example, and is therefore useful as a model system for study of DNA bending.

Bending locus in K-DNA fragments

Quantitative measurement of the relative gel mobility of a set of circularly permuted fragments allows us to extrapolate to the position of the cut which would yield maximum gel mobility, and thereby locate the centre of the molecular bend. This procedure is illustrated in Figs 2 and 3, which refer respectively to the 423 bp K-DNA fragment, and a 241 bp subfragment corresponding to the left end of the K fragment. In each case, a cloned tandem dimer of the fragment was cleaved with a variety of enzymes which cut once in the sequence; we observe a striking variation of gel mobility with position of the end. These plots utilize the corrected sequence of the K-DNA fragment as determined by S. Dieckmann and J. C. Wang and by P. T. Englund and J. C. Marini (personal communications).

Extrapolation of the nearly linear portions of the graph of gel mobility against position of the molecular ends yields an estimate of bp 148 ± 5 for the bending locus in the K-DNA fragment (Fig. 2), with an estimate of bp 140 ± 3 for the 241-bp subfragment in Fig. 3. The difference between these values is not due to experimental error, but reflects the complex shape of the larger K-DNA fragment, which may contain another significant bend in the right hand half of the molecule. In general,

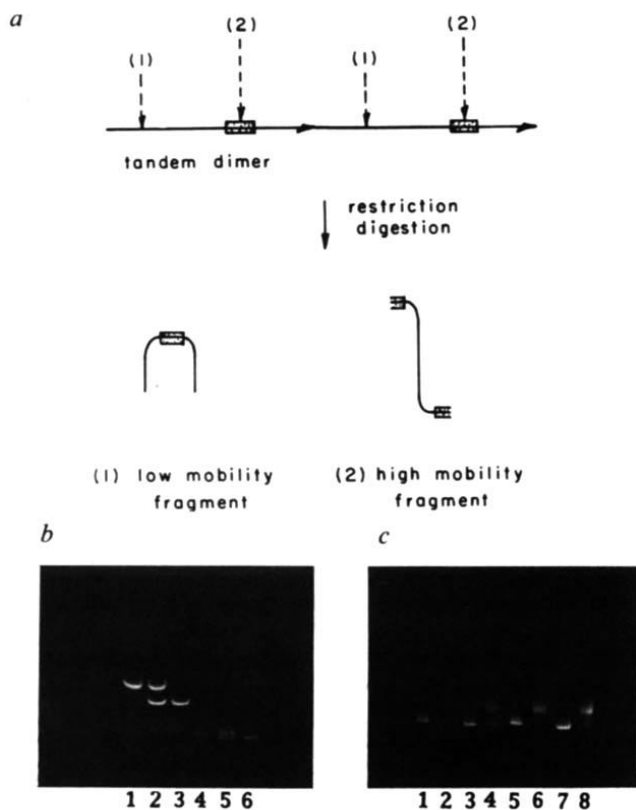


Fig. 1 *a*, Construction of circularly permuted variants of a DNA sequence differing in conformation because of the location of a bending locus shown by the box. The tandem dimer is cleaved with two different enzymes (1) and (2), yielding fragment (1), which is strongly bent and therefore of low gel mobility, and fragment (2), which is nearly linear since the bend is near the end, and therefore of high mobility. *b*, Position effect of the bending site. Plasmid pJW423 (constructed by S. Levene and J. C. Wang) is a plasmid containing the Sau3A K-DNA fragment at the *Bam*HI site of vector pJW37. Plasmid pJW37 (constructed by J. C. Wang) is a derivative of pBR322 with deletion of the region from nucleotide 236 to 2,352, and with the *Hind*III site replaced with a *Bam*HI site flanked by two *Hind*III sites. pJW423 and pJW37 DNA were linearized with different restriction enzymes which cut only once in the plasmids, either adjacent to (*Bam*HI) or far from (*Pst*I) the K-DNA insert. The digests were loaded onto a 4% polyacrylamide gel with TBE buffer (89 mM Tris, 89 mM boric acid, 2.5 mM EDTA). Samples in each lane are 1, pJW423-*Pst*I; 2, pJW423-*Pst*I + pJW423-*Bam*HI; 3, pJW423-*Bam*HI; 4, pJW37-*Pst*I; 5, pJW37-*Pst*I + pJW37-*Bam*HI; 6, pJW37-*Bam*HI. *c*, Position effect of protein-induced bending. Plasmid pHW101 is a plasmid containing a single copy of 203 bp *lac* promoter DNA at the *Eco*RI site of pHW1 (a derivative of pJW37 with deletion of the *Bam*HI site). pHW101 was linearized with *Hind*III or *Pst*I enzyme. The linear plasmid DNA was incubated with CAP (kindly provided by H. N. Liu-Johnson and J. Rice) plus 10 μ M cyclic AMP, or with *lac* repressor (a gift from M. Leahy) (1:1 molar ratio of protein to DNA) for 1 h and loaded directly onto a 4% polyacrylamide gel in TBE buffer with 10 μ M cyclic AMP. Samples in each lane are pHW101-*Pst*I (1 and 5), pHW101-*Pst*I + CAP (2), pHW101-*Hind*III (3 and 7), pHW101-*Hind*III + CAP (4), pHW101-*Pst*I + *lac* repressor (6), and pHW101-*Hind*III + *lac* repressor (8).

if there is only one major bending locus in a fragment, the mobility curve should be symmetrical and roughly sinusoidal with period equal to the length of the DNA fragment. An asymmetric variation of gel mobility with end position is an indication of more complicated shape. The mobility plot for the 241 bp subfragment is more regular than that for the 423 bp fragment, with a symmetric variation of gel mobility about a maximum and a minimum. The maximum (bp 140) and the minimum (bp 20) of the mobility curve are separated by 120 base pairs, or half the fragment length, as expected. (Note that mobility decreases in moving up the vertical axis in Figs 2-4.)

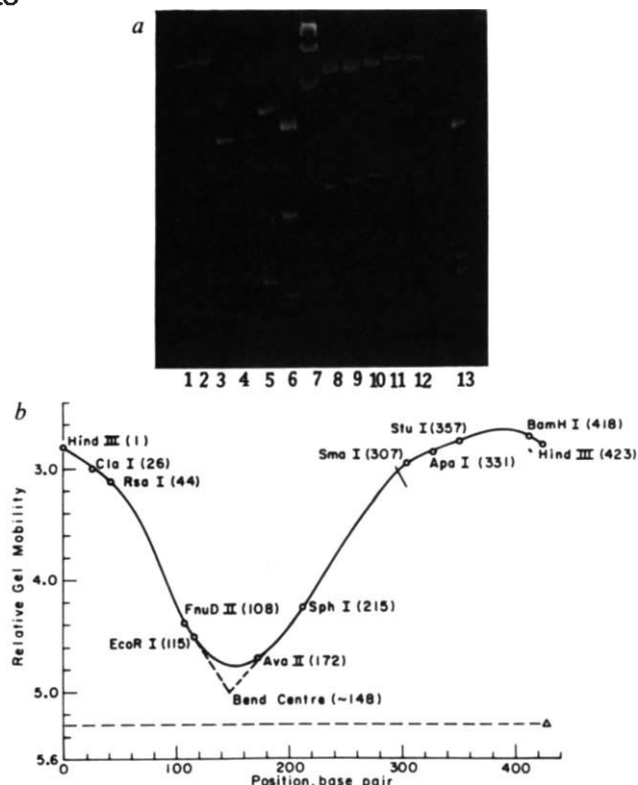


Fig. 2 *a*, 5% Polyacrylamide gel electrophoresis (TBE) of circularly permuted K-DNA fragments. K-DNA fragments were self-ligated and cloned into the *Hind*III site of the vector pHW1. Clones with tandem repeats of K-DNA fragments were identified through a boiling minilysis procedure⁴⁰. pHW122 containing a tandem dimer of the K-DNA insert was digested with different restriction enzymes which cut K-DNA only once. The digests were loaded directly onto the gel. Samples in each lane are 1, *Hind*III; 2, *Cl*aI; 3, *R*saI; 4, *F*nuDII; 5, *E*coRI; 6, *A*vaII; 7, *S*phI; 8, *S*maI; 9, *A*paI; 10, *S*tuI; 11, *B*amHI; 12, *H*indIII (all restriction enzymes purchased from New England Biolabs). The highest mobility band of each lane is the circularly permuted K-DNA fragment; other fragments arise from the vector. Circularly permuted *F*nuDII K-DNA fragment shown in lane 4 was purified fragment, because of the complexity of the *F*nuDII digestion pattern. Lane 13 is pBR322 *H*inI as a size marker. *b*, Mapping of the bending site on the K-DNA fragment. The relative gel mobilities of circularly permuted K-DNA fragments were plotted against the position of the restriction sites in the K-DNA fragment. The number in parenthesis at each restriction site is the distance in base pairs from the left *Hind*III site of the fragment to the position of the cut on the 5' to 3' strand. The bend centre was estimated through extrapolation to be at 148 ± 5 bp from the left end of the fragment. $\circ$, Relative mobilities of circularly permuted K-DNA fragment; Δ , relative mobility of normal 423 bp.

DNA sequence at K-DNA bending locus

The DNA sequence⁶ around the bending locus of K-DNA is shown in Fig. 3*b*. The striking pattern contains periodically repeated (dA)₅₋₆ tracts separated by four to six base pairs of G+C rich sequences. The centre of bending (bp 140) for the sub K-DNA fragment is positioned in the middle of this DNA sequence pattern. There are other oligo(dA) tracts scattered throughout the K-DNA sequence but they are not as regularly distributed as in the region of the bending centre. It is known that poly(dA)·poly(dT) exhibits distinctly different characteristics from other members of the B DNA family: it has a helical repeat of 10 bp instead of 10.5^{31,32} and does not form nucleosomes upon binding with histones³³⁻³⁵. The unique sequence of oligo(dA) tracts and their striking distribution around the observed bending centre lead us to reaffirm the periodic pattern of oligo(dA) tracts as the cause of the bending⁶. We note in addition that each oligo(dA) tract is bordered by C and T at its 5' and 3' ends respectively. Hence we conclude that CA₅₋₆T tracts,

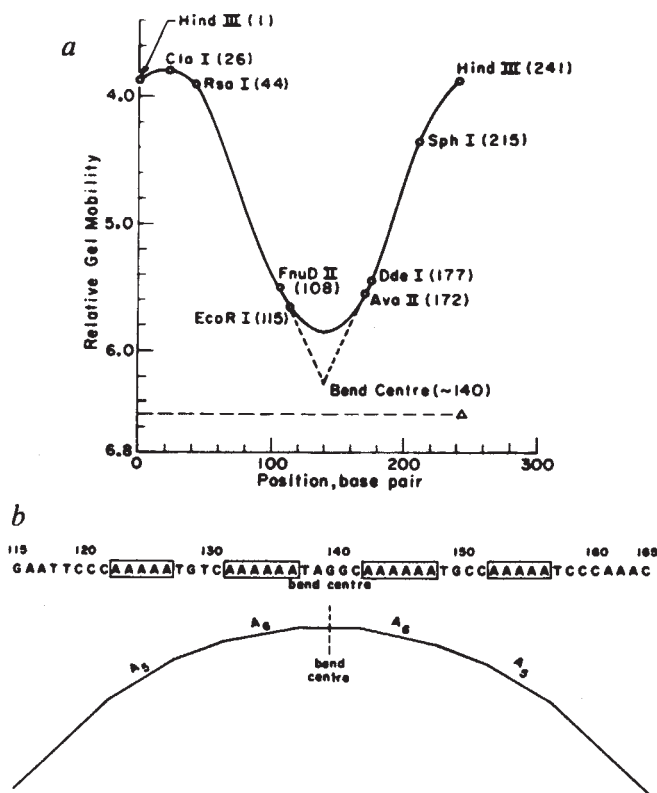


Fig. 3 *a*, Mapping of the bending site on a sub K-DNA fragment. Plasmid pJW423 DNA was digested with *Hae*III enzyme. The *Hae*III fragment containing the 237 bp *Hind*III-*Hae*III region of K-DNA was purified and *Hind*III linkers (New England Biolabs) were attached. After digesting with *Hind*III enzyme, the sub K-fragment with a *Hind*III site at each end was purified and self-ligated, then cloned into the *Hind*III site of pHW1. pHW132 containing a tandem dimer of the sub K-DNA insert was digested with different enzymes and loaded directly onto an 8% polyacrylamide gel in TBE buffer. The relative gel mobilities of circularly permuted sub K-DNA fragments were plotted against the position of the restriction sites of the sub K-DNA fragment. From this experiment we estimate the bend centre to be at 140 ± 3 bp from the left end of the fragment. $\circ$, Relative mobilities of circularly permuted sub K-DNA fragments; Δ , relative mobility of control 241 bp DNA fragment. *b*, DNA sequence⁶ and junction bending model of K-DNA around the bending locus of sub K-DNA⁶. Note the periodic repetition of the sequence-CA₅₋₇T. Differences in helical geometry between dA·dT tracts and adjacent DNA segments are proposed as the origin of junction bends. The figure shows a planar projection of the bending angles which result when a 12° junction bending angle is assumed between dA·dT tracts and the adjacent B-DNA helices. Calculations used the rotational matrix described by Flory⁴³, and assumed a helical twist of 34.3° per base pair for B-DNA segments, and 36.0° per base pair for dA·dT tracts^{31,32}.

periodically repeated at 10 bp intervals, are responsible for sequence-directed bending of the K-DNA fragment.

DNA bending induced by binding of CAP

Figure 1c shows the position effect of the CAP binding site on the gel mobilities of CAP-plasmid DNA complexes. Circular plasmids containing the *lac* promoter sequence were linearized respectively by two different restriction enzymes which cleaved only once in the plasmid. The linear plasmids generated have the intact *lac* promoter sequence either in the centre (*Pst*I digestion) or near the end (*Hind*III digestion) of the DNA fragment. The complex of CAP with *Pst*I-cleaved linear plasmid shows significantly slower mobility than that of CAP with the *Hind*III cleaved plasmid. In contrast, the complexes of *lac* repressor with these two linear plasmids show no detectable variation of mobility with *lac* promoter location. This indicates that the variation of mobility for the CAP-plasmid complex is

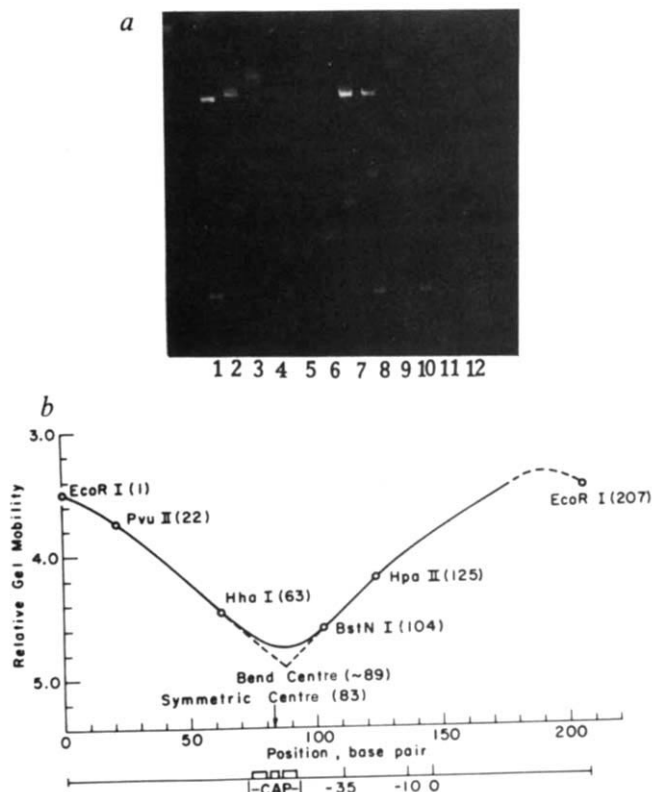


Fig. 4 *a*, 5% polyacrylamide gel electrophoresis (TBE) of the complexes of CAP and a circularly permuted DNA fragment which contains the *E. coli lac* promoter region. DNA fragments of 203 bp containing the *lac* control region were self-ligated and cloned into the *Eco*RI site of vector pHW1. The clones with tandemly repeated 203 bp DNA inserts were identified through a boiling minilysis procedure. pHW104 DNA, which contains four tandem repeats of the 203 bp. DNA, was cut with different restriction enzymes to generate circularly permuted fragments. CAP or *lac* repressor and circularly permuted *lac* DNA fragments were incubated and loaded directly onto the gel as described previously³⁷. Samples in each lane are: 1, pHW104-*Eco*RI; 2, pHW104-*Eco*RI+CAP; 3, pHW104-*Pvu*II+CAP; 4, purified *Hha*I circularly permuted fragment+CAP; 5, purified *Bst*NI fragment+CAP; 6, purified *Hpa*II fragment+CAP; 7, pHW104-*Eco*RI+CAP; 8, pHW104-*Eco*RI+*lac* repressor; 9, purified *Pvu*II fragment+*lac* repressor; 10, purified *Hha*I fragment+*lac* repressor; 11, purified *Bst*NI fragment+*lac* repressor; 12, purified *Hpa*II fragment+*lac* repressor. Purified fragments were used in some cases because of the complexity of the restriction pattern, but no difference in gel mobility of protein-DNA complexes was observed when whole digests were used. *b*, Mapping of the bending site in the CAP-203 bp DNA complex. The relative gel mobility of the complex of CAP and circularly permuted 203-bp *lac* DNA fragments is plotted against the position of restriction sites in the DNA fragment, expressed as the distance along the 5'→3' (upper) strand from the left hand *Eco*RI site in the normal orientation of the 203-bp fragment. The bend centre was estimated through extrapolation at 89 ± 2 bp. The bottom line shows the functional map of the 203-bp *lac* promoter fragment, with twofold symmetric regions of the CAP binding sequence indicated by boxes. The bend centre is significantly displaced towards the promoter from the centre of the protein binding site (83 in this coordinate frame). The expected minimum in gel mobility is shown as the dotted region of the curve at its expected position of 191 bp, displaced by half the fragment length, or 103 bp, from the mobility maximum at 89 bp.

not simply the result of CAP location in the linear plasmid but rather reflect the conformation of DNA in the complex. By analogy with the observations on the K-DNA plasmid (Fig. 1b), these results strongly suggest that CAP bends *lac* promoter DNA, but *lac* repressor does not do so to an extent detectable by this assay.

Although the molecular weight of CAP dimer is much smaller than *lac* repressor tetramer, as shown in Fig. 4a, the mobility

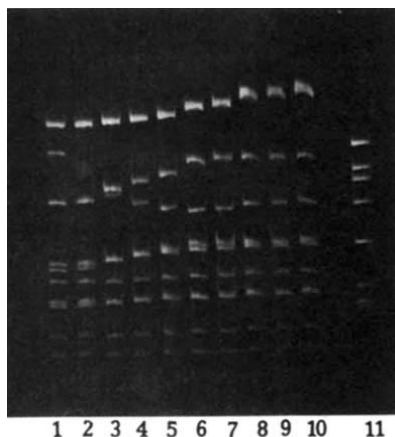


Fig. 5 5% gel electrophoresis titration of *Sau*3A restriction fragments of plasmid pPE103⁶ (pBR322 with the *Sau*3A K-DNA fragment cloned into the *Bam*HI site) with distamycin. To aliquots of 1 μ g restriction fragments were added various amounts of distamycin, and the samples were loaded onto the gel directly. r_{add} of distamycin from left to right: 1, 0; 2, 0.005; 3, 0.01; 4, 0.03; 5, 0.05; 6, 0.07; 7, 0.09; 8, 0.15; 9, 0.2; 10, 0.3 per base pair. The second DNA fragment from the top in lane (1) is the cloned kinetoplast DNA; other fragments are generated from the vector pBR322 used in the pPE103 clone. As added distamycin was increased the mobility of K-DNA increased and that of the third fragment from the top decreased. They crossed at $r_{add} = 0.01$ and K-DNA became the third fragment from the top thereafter. Lane 11 is pBR322-*Alu*I as a size marker.

of the CAP-203 DNA complex is not much faster than that of *lac* repressor complexed with the same DNA fragment. When DNA length is increased from 203 bp to 2,400 bp, the mobility of the complex of CAP and *Pst*I cleaved plasmid becomes significantly slower than that of *lac* repressor complexed with the same linear plasmid. No difference in the relative gel mobilities for the complexes could be detected with a variation of electrophoretic buffer from 13 mM to 100 mM ionic strength (data not shown). All of these findings strongly indicate that the different gel mobilities of the complexes are not the consequence of differences in the charge or molecular weight of the proteins but of the DNA conformations in the complexes. The size of the DNA fragment apparently also influences the extent of reduction in gel mobility resulting from bending. Keeping the bending locus in the centre of the fragment, we found that, for K-DNA, the longer the fragment, the greater the percentage increase in apparent molecular weight from gel mobility (data not shown). It could be this length effect which causes a greater mobility retardation of CAP compared to *lac* repressor, both complexed with a large DNA fragment.

Asymmetric bending of *lac* promoter DNA

Figure 4 shows the result of electrophoretic analysis of circularly permuted *lac* promoter fragments interacting with CAP or *lac* repressor. Only the CAP-DNA complex exhibits a strong dependence of gel mobility on the position of the molecular ends, implying that neither the DNA fragment alone nor the repressor-DNA complex is appreciably bent.

Extrapolation of the gel mobility data to the maximum leads us to identify bp 88–90 (counting from the left end of the fragment) as the centre of the bending locus; in this coordinate frame the centre of the CAP binding site is at 83. Several repetitions of the experiment encourage us to set rather narrow error limits of ± 2 bp on the position estimate. The bending is not symmetric about the centre of the CAP site, as is shown by the significant difference in gel mobility for cuts by *Hha*I and *Bst*NI, which are nearly equally distant (20 and 21 bp) from the centre of the CAP site. Since the individual strand scissions produced by these two enzymes are staggered by only 2 and 1 bp respectively, it is unlikely that differences in the resulting DNA end structures are responsible for the apparent shift of

the centre of bending symmetry away from the centre of the protein binding site by an estimated 5–7 bp.

Asymmetry of the CAP-promoter complex is also indicated by the observation that only one of the protein subunits binds cyclic AMP tightly^{36,37}. Asymmetry is, of course, implied by the fact that the promoter is found on only one side of the CAP site. Unfortunately there are no known restriction sites near the expected minimum in gel mobility (188 bp), shown at its predicted position by the dotted line on the graph in Fig. 4b. In other cases it may be possible to define the bending locus equally well from the gel mobility minimum.

Structural asymmetry is also found in the crystal structure of the CAP dimer²¹. The relative orientation of the two structural domains, DNA binding and cyclic AMP binding, are different in the two subunits of the CAP dimer. In one case, the two domains are tightly packed to form a close conformation, while in the other subunit, the two domains form a more open cleft. The two DNA binding domains in the dimer are not parallel, but rather are oriented at an angle as large as 28°²¹. This structural asymmetry in the CAP dimer could help induce the asymmetric DNA bending as we observed by the gel electrophoresis method. Moreover, bending right-handed B-DNA in the CAP-DNA complex could increase the number of favourable protein-DNA contacts (T. Steitz, personal communication).

Removal of DNA bends by distamycin

Figure 5 shows that small amounts of the oligopeptide antibiotic distamycin normalize the gel mobility of the K-DNA fragment. In the presence of one distamycin molecule added per 30 bp, the mobility of the fragment is similar to that of a 423 bp DNA complexed with distamycin. The mobility changes seen in Fig. 5 for other fragments arising from the vector probably reflect induced bends, and are addressed in elsewhere⁵.

We found that distamycin at low drug:DNA ratios is also able to increase the mobility of the CAP-DNA complex without displacing CAP from the DNA, as judged from the sharp electrophoretic band found for the CAP-DNA complex in the presence of drug. In addition, distamycin greatly reduces the mobility difference between the circularly permuted variants of the complex (data not shown). We draw the tentative conclusion that distamycin can reduce bending in the CAP-promoter DNA complex. The similarities between this result and the observations on the K-DNA fragment suggest the intriguing possibility that bending may share a common structural basis in the two cases.

Models for DNA bending and activation

Models for DNA bending in chromatin have generally assumed either smooth bending or kinks¹; these two basic possibilities for the CAP-DNA complex are shown in diagrammatic form in Fig. 6. Smooth bending (Fig. 6a) involves small changes in helix direction at each base pair, a possibility which we cannot presently exclude. However, we prefer the kinked model (Fig. 6b), which we call junction bending, because it leads to more interesting hypotheses as a focus for further work. Specifically, we propose that in both K-DNA and the CAP-DNA complex, bending occurs at the junctions between regions of B-DNA and sections containing an altered helical conformation. For

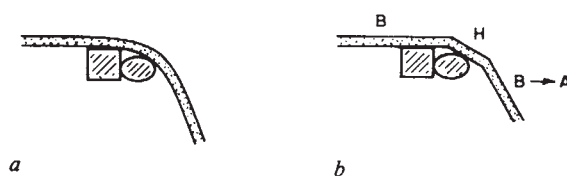


Fig. 6 Models for DNA bending by CAP protein. *a*, Smooth bending; *b*, junction bending. The protein subunits are shown as non-equivalent because only a single cAMP is taken up in formation of the specific CAP-promoter complex.

example, Selsing *et al.*⁸ and Arnott *et al.*⁹ have shown from model building that a helix bend is expected at the junction of B and A DNA regions.

In the case of K-DNA, we propose that oligo dA tracts have an altered helix structure, perhaps like the heteronomous (H) DNA structure devised by Arnott *et al.*³⁹ to explain the fibre diffraction data for poly(dA)·poly(dT). When these tracts are about 5 bp (half a turn) long, the junction bends will point in the same direction, yielding a net molecular bend, as shown in Fig. 3b. (Junction bends spaced by a full helical turn yield only a helix axis dislocation, not a net bend.) If these 5-bp tracts are separated by 5-bp spacers, then the bending will be additive and in the same plane. For example, with a junction bend of about 12° for the H-B DNA junction, half the angle suggested by Arnott *et al.*⁷ for the A-B junction, then the projection of the DNA helix on the plane around the bending locus of K-DNA can be drawn as shown in Fig. 3c. Eight nearly inphase 12° bends generate an overall bend of about 90°, with the centre of bending located around bp 140, in close agreement with the results determined by the gel method for the sub-K fragment. This model readily explains the effect of distamycin on K-DNA, since it is known that the drug prefers the B helical form³⁸. Reversion of dA tracts to B helix would eliminate junction bends; the same phenomenon may explain the disappearance

of electrophoretic anomalies at elevated temperature⁴¹. Work is under way in our laboratory to explore quantitative application of this model to the K-DNA structure⁴².

By analogy with the model for K-DNA bending, we propose that CAP may bend promoter DNA by induction of a structure such as the H form, creating junction bends. This proposal offers several possibilities for an influence of CAP binding on RNA polymerase-promoter interactions. For example, polymerase may recognize the H conformation directly, sliding from there to its normal interaction sites at -35 and -10. Alternatively, the H conformation may potentiate transient conversion of the adjacent promoter B-DNA to a conformation such as A which more readily binds polymerase. Finally, it is also possible that the role of DNA bending is primarily to create CAP-polymerase interactions which would not be sterically possible in a linear promoter. Further work will be required to sort through these possibilities.

This work was supported by NIH grant GM21966 and grant NP247 from the ACS. We thank Professor James Wang for sending us the nucleotide sequence of the K-DNA fragment in advance of publication and for help on the construction of pJW423, and Professor Struther Arnott for discussions on the nature of helix-helix junctions. H.M.W. thanks Dr Alice Cheung for advice on cloning techniques.

Received 22 September 1983; accepted 6 January 1984.

- McGhee, J. D. & Felsenfeld, G. A. *Rev. Biochem.* **49**, 1115-1156 (1980).
- Klevan, L. & Wang, J. C. *Biochemistry* **19**, 5229-5234 (1980).
- Better, M., Lu, C., Williams, R. C. & Echols, H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5837-5841 (1982).
- Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthews, B. W. *Nature* **298**, 718-723 (1982).
- Wu, H. M., Dattagupta, N. & Crothers, D. M. *Proc. natn. Acad. Sci. U.S.A.* **80** (in the press).
- Marini, J., Levene, S., Crothers, D. M. & Englund, P. T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7664-7668 (1982).
- Selsing, E. & Wells, R. J. *biol. Chem.* **254**, 5410-5416 (1979).
- Selsing, E., Well, R., Alden, C. & Arnott, S. J. *biol. Chem.* **254**, 5417-5422 (1979).
- Arnott, S. *et al. Cold Spring Harb. Symp. quant. Biol.* **47**, 53-65 (1983).
- de Crombrughe, B. & Pastan, I. in *The Operon* (eds Miller, J. H. & Reznikoff, W. S.) 303-324 (Cold Spring Harbor Laboratory, New York, 1978).
- Ullmann, A. & Danchin, A. *Adv. Cyclic Nucleotide Res.* **15**, 1-53 (1983).
- Majors, J. *Nature* **256**, 672-674 (1975).
- Gilbert, W. in *RNA Polymerase* (eds Losick, R. & Chamberlin, M.) 193-205 (Cold Spring Harbor Laboratory, New York, 1976).
- Simpson, R. B. *Nucleic Acid Res.* **8**, 759-766 (1980).
- Schmitz, A. *Nucleic Acid Res.* **9**, 277-292 (1981).
- Caruthers, M. H. *et al. Cold Spring Harb. Symp. quant. Biol.* **47**, 411-418 (1983).
- Dickson, R. C., Abelson, J. N., Barnes, W. M. & Reznikoff, W. S. *Science* **182**, 27-31 (1975).
- Nakanishi, S., Adhya, S., Gottesman, M. & Pastan, I. *J. biol. Chem.* **254**, 4050-4056 (1974).
- Nakanishi, S., Adhya, S., Gottesman, M. & Pastan, I. *J. biol. Chem.* **250**, 8202-8208 (1975).

- Wells, R. D. *et al. Proc. Nucleic Acid Res. molec. Biol.* **24**, 167-267 (1980).
- McKay, D. B. & Steitz, T. A. *Nature* **290**, 744-749 (1981).
- Ebright, R. H. & Wong, J. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4011-4015 (1981).
- Taniguchi, T., O'Neill, M. & de Crombrughe, B. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5090-5094 (1979).
- Lee, N. L., Gielow, W. O. & Wallace, R. G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 752-756 (1981).
- Valentin-Hansen, P. *EMBO J.* **1**, 1049-1054 (1982).
- Kolb, A. & Buc, H. *Nucleic Acids Res.* **10**, 473-485 (1982).
- Unger, B., Clore, M., Gronenborn, A. & Hillen, W. *EMBO J.* **2**, 289-293 (1983).
- Cooper, T. B. in *The Tools of Biochemistry* 194-253 (Wiley, New York, 1977).
- Lerman, L. S. & Frisch, H. L. *Biopolymers* **21**, 995-997 (1982).
- Lumpkin, O. J. & Zimm, B. H. *Biopolymers* **21**, 2315-2316 (1982).
- Peck, L. J. & Wang, J. C. *Nature* **292**, 375-378 (1981).
- Rhodes, D. & Klug, A. *Nature* **286**, 573-578 (1980).
- Simpson, R. T. & Kunzler, P. *Nucleic Acids Res.* **6**, 1387-1415 (1979).
- Rhodes, D. *Nucleic Acids Res.* **6**, 1805-1816 (1979).
- Kunkel, G. R. & Martinson, H. G. *Nucleic Acids Res.* **9**, 6869-6888 (1981).
- Garner, M. M. & Revzin, A. *Biochemistry* **24**, 6032-6036 (1982).
- Fried, M. G. & Crothers, D. M. *Nucleic Acid Res.* **11**, 141-158 (1983).
- Wartell, R., Larson, J. & Wells, R. J. *J. biol. Chem.* **249**, 6719-6731 (1974).
- Arnott, S., Chandrasekaran, R., Hall, I. & Puigjaner, L. *Nucleic Acid Res.* **11**, 4141-4155 (1983).
- Holmes, D. S. & Quigley, M. *Analyt. Biochem.* **114**, 193-197 (1981).
- Wu, H. M., thesis, Yale Univ. (1982).
- Levene, S. & Crothers, D. M. *J. Biomolecular Structure Dynamics* **1**, 429-435 (1983).
- Flory, P. J. *Statistical Mechanics of Chain Molecules* (Wiley, New York, 1969).

Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-II_A gene

Michael Karin*, Alois Haslinger*, Heidi Holtgreve*, Robert I. Richards†, Paul Krauter‡, Hannes M. Westphal‡ & Miguel Beato*

*Department of Microbiology, University of Southern California School of Medicine, Los Angeles, California 90033, USA

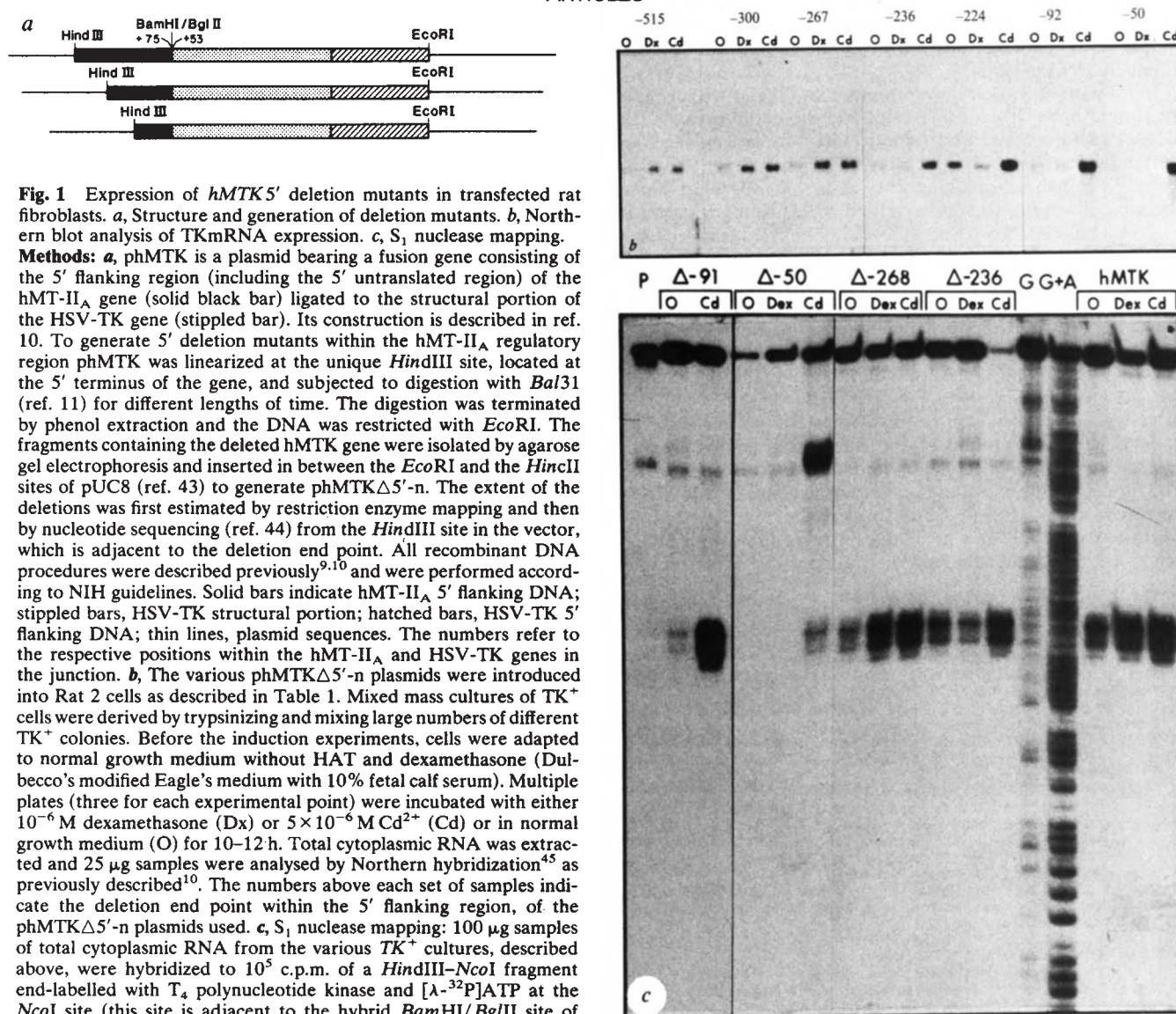
† Department of Genetics, Research School of Biological Sciences, Australian National University, Canberra ACT 2601, Australia

‡ Physiological Chemistry Institute, Philipps University, 3550 Marburg, West Germany

Deletion experiments have defined two stretches of DNA (genetic elements), lying close to the promoter for a human gene for metallothionein, that separately mediate the induction of the gene by heavy metal ions, particularly cadmium, and by glucocorticoid hormones. The element responsible for induction by cadmium is duplicated, yet a single copy is fully functional; the element responsible for induction by glucocorticoid hormones is coincident with the DNA-binding site for the glucocorticoid hormone receptor.

METALLOTHIONEIN (MT) genes, encoding low molecular weight heavy-metal binding proteins, are a useful experimental system for studying transcriptional regulation of higher eukaryotic genes. In mammals, MT genes are induced in a variety of cell lines and tissues by administration of either heavy metal ions or glucocorticoid hormones¹⁻⁴. Both inducers

are primary inducers of MT mRNA⁵, acting at the transcriptional level^{4,6}. The kinetic analysis of MT mRNA induction in HeLa cells, and the differential induction response to metal ions and glucocorticoids after inhibition of protein synthesis, suggest that the two classes of inducers act by independent mechanisms^{5,7,8}.



We recently introduced the cloned human MT-II_A gene (hMT-II_A) (ref. 9) into Rat 2 fibroblasts and found that it retains its responsiveness to both glucocorticoids and heavy metals. Also, a fusion gene (hMTK), constructed by joining the 5' flanking region of hMT-II_A gene to the structural portion of the herpes simplex virus (HSV) thymidine kinase (TK) gene, is responsive to both inducers. The induction of both the intact hMT-II_A gene and of the hMTK fusion gene is due to increased transcription following either Cd²⁺ or dexamethasone administration¹⁰.

To map the regulatory DNA elements and determine whether they serve as binding sites for various regulatory proteins we constructed progressive deletion mutants in the 5' flanking region of the hMTK gene and tested them for both hormonal and heavy metal induction *in vivo*, after transfection into rat fibroblasts, and for their ability to bind to the rat liver glucocorticoid hormone-receptor complex *in vitro*.

We find at least three separate regulatory elements within the 5' flanking region of the hMT-II_A gene, one, present in two copies, is involved in heavy metal ion induction. The second

element centred around nucleotides -70 to -90 is important for the maintenance of the basal level of expression. The third element, centred 250–260 base pairs (bp) upstream to the start site of the transcription, is involved in hormonal induction. This element is also the binding site of the purified glucocorticoid hormone-receptor complex to the human gene.

Two upstream regulatory elements

We constructed a chimaeric gene consisting of 800 bp of the 5' flanking region, containing the promoter elements, the cap site and the 5' untranslated region of the hMT-II_A, fused to the transcribed portion of the HSV TK gene. When this fusion gene (hMTK) is transcribed from the hMT-II_A promoter it generates a functional TK mRNA at levels increased by treatment of transfected cells with either dexamethasone or Cd²⁺. Both inducers act by increasing the transcription rate of the transfected gene¹⁰. We used the hMTK fusion for the construction of deletion mutants to characterize the regulatory elements of the hMT-II_A promoter. To generate a series of 5' deletions, phMTK was linearized at the unique *Hind*III site, at its 5' terminus, and

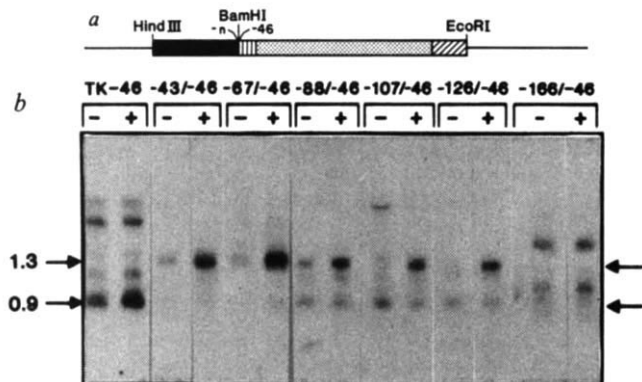


Fig. 2 Analysis of 3' deletion mutants. *a*, Construction of 3' deletions and hybrid promoters. The diagram illustrates the structure of the pHMK-n/-46 fusions: solid bar, hMT-II_A 5' flanking DNA; stippled bar, HSV-TK transcribed region; hatched bars, HSV-TK 5' and 3' flanking DNA; thin lines, plasmid DNA. -n1, deletion end point within the hMT-II_A gene; -46, deletion end point within TK gene. *b*, Expression of the pHMT-n/-46 fusions. Arrows refer to 1.3 and 0.9-kb transcripts of HSV-TK gene, plus and minus signs denote presence or absence of Cd²⁺ before RNA extraction (see below).

Methods: *a*, To generate 3' deletions in the hMT-II_A gene 5' flanking regions, p84H, a HindIII subclone of hMT-II_A gene⁹, was linearized at the single NcoI site (at the AUG initiator codon of the gene). The linearized DNA was digested for different times with exonuclease Bal31 (New England BioLabs). After terminating the digestion, the DNA was restricted with HindIII, and fragments of the appropriate size containing the hMT-II_A gene 5' flanking region were isolated on agarose gels and subcloned into the HindIII and HincII sites of pUC8 (ref. 43) to generate pHMT-II-Δ3'-n, whereas n stands for the deletion end point within the hMT-II_A 5' flanking region determined by chemical sequencing⁴⁴. Plasmids pTK-Δ5'-46 and pTK-Δ5'-109 (ref. 13) were digested with EcoRI (a partial digest in the case of pTK-Δ5'-109) and BamHI, and the fragments containing the HSV-TK gene 5' deletion mutants were isolated and cloned between the BamHI and EcoRI sites of pHMT-II-Δ3'-n to generate pHMT-n/-46 and pHMT-n/-109 respectively. There are 11 nucleotides (5'-GACGGATCCGG-3') derived from the vector and the Bam linker used by McKnight that separate the HSV-TK and the MT-II gene sequences. *b*, The various pHMK-n/-46 plasmids were introduced into Rat 2 cells as described in Table 1. Mixed mass cultures were derived and treated either with (+) or without (-) 5 × 10⁻⁶ Cd²⁺ before extraction of total cytoplasmic RNA. 25-μg RNA samples were analysed by Northern hybridization⁴⁵. TK-46, parent TK gene deletion mutant used for construction of the fusions. The numbers above each set of samples denote the deletion end point within the hMT-II_A regulatory region (-n1) and the deletion end point of the HSV-TK gene used (-46). Arrows, 1.3 and 0.9 kb transcripts of the HSV-TK gene.

digested with the processing exonuclease Bal31¹¹. Different deletion derivatives were isolated by gel electrophoresis and recombined (see Fig. 1*a* for details). The progressive deletions in the 5' to 3' direction are referred to as pHMTKΔ5'-n where n is the nucleotide present at the deletion end point within the hMT-II_A 5' flanking region (see Fig. 1*a*). The deletion mutants were introduced into Rat 2 cells by transfection and selection in hypoxanthine-aminopterin-thymidine (HAT) medium. In the past, we have shown that selection with dexamethasone leads to 5–10-fold increase in the number of TK⁺ clones generated by pHMTK¹⁰, due to increased transcriptional activity of the marker gene. This was used as a rapid assay to localize the element involved in hormonal induction. The results of such transfection experiments are shown in Table 1. Also the comparison of transfection efficiencies of the pHMTKΔ5' mutants in the absence of hormone, helps to map regulatory elements affecting the basal level of expression.

With dexamethasone, therefore, transformation efficiency of hMTKΔ5' mutants, containing at least 268 bp of the 5' flanking region, is increased 5- to 14-fold. Constructs containing between 91 to 236 bp of the hMT-II_A regulatory region while retaining similar basal activity are no longer hormone responsive (Table

Table 1 Effect of dexamethasone on transformation efficiencies of hMTK deletion mutants

Deletion end points	No. of TK ⁺ colonies		Fold increase
	-dexamethasone	+dexamethasone	
Expt 1			
-585	27 ± 2	137 ± 4	5.1
-515	35 ± 4	166 ± 4	4.7
-445	26 ± 1	118 ± 13	4.5
-300	36 ± 3	326 ± 23	9.1
HSV-TK	238 ± 7	253 ± 11	1.06
Expt 2			
-300	28 ± 1	220 ± 3	8
-236	18 ± 1	33 ± 1	1.8
-223	31 ± 2	28 ± 4	0.9
-74	2 ± 0	3 ± 1	1.5
HSV-TK	180 ± 6	175 ± 7	0.97
Expt 3			
-300	22 ± 3	114 ± 1	5.2
-268	6 ± 0	87 ± 19	14.5
-91	15 ± 3	20 ± 3	1.3
-50	1 ± 0.5	2 ± 0.5	2.0
HSV-TK	110 ± 5	140 ± 20	1.3

Rat cells were transfected with a calcium phosphate coprecipitate of pHMTKΔ5'-n or pHSV-TK (2 μg per 100 mm plate) and calf thymus carrier DNA (10 μg per 100 mm plate). Transfection and selection procedures were as previously described¹⁰. After 2 weeks of growth in selective medium either in the absence or in the presence of dexamethasone (10⁻⁶M) the colonies were fixed, stained and counted¹⁰. Numbers represent averages of two or three plates.

1). Deletion mutants containing less than 91 bp of the 5' regulatory region of the hMT-II_A gene exhibit markedly reduced transformation efficiency.

To examine the effect of inducers on expression of the hMTK deletions, more directly, we treated stable TK⁺ cultures, derived from many different individual clones with either dexamethasone or Cd²⁺ for 10 h, extracted total cytoplasmic RNA and determined the relative level of TK mRNA present. The average activity of a given mutant can be determined by examining its expression in a mixed population of cells. Individual clones vary in their levels of expression and are not suitable for such an analysis. The level of TK mRNA was increased by both dexamethasone and Cd²⁺ in all of the TK⁺ cultures derived from transfection with hMTK 5' deletion mutants containing 268 bp or more of the hMT-II_A 5' regulatory region (Fig. 1*b*). Cultures derived from deletion mutants containing 236 bp or less of 5' flanking sequences no longer respond to the hormone, but are still metal inducible. The basal level of expression of the mutants did not vary much in most of the cultures tested, except for pHMTKΔ5'-50, in which the basal level of activity of the hMT-II_A promoter is essentially undetectable, however, it is still responsive to Cd²⁺. Thus, the basal level of hMTK mRNA seems to correspond quite well to the relative transfection efficiencies of the various mutants determined in the absence of dexamethasone (see Table 1). The copy number of pHMTKΔ5'-n DNA integrated into the TK⁺ cells was examined and found to vary little (not more than twofold) amongst the different cultures (M.K. and A. Heguy, unpublished results.)

Next we investigated whether the deletion of 5' flanking sequences has any effects on selection of start sites for transcription of the hMT-II_A promoter. Using the S₁ nuclease mapping technique¹² we find that none of the deletions effects the choice of initiation sites for the hMT-II_A promoter (Fig. 1*c*). Induction by either Cd²⁺ or dexamethasone increases the frequency of initiation at those sites.

Heavy-metal ion responsive element

Analysis of the 5' deletion mutants indicates that only 50 nucleotides of the 5' flanking region are necessary for heavy-metal ion induction. Our method of analysis depends on transcriptional activity for obtaining transformants and we were not successful in analysing any further deletion mutants which all fell within

the TATA box region. To achieve a better definition of the heavy-metal ion responsive element we have used a new approach to mapping of regulatory elements. To map the 3' boundaries of such elements, *Bal*31 deletions were initiated from a unique restriction site (*Nco*I) 3' to the promoter. In the course of mutagenesis the proximal promoter elements (TATA box and cap site) are deleted; therefore, to restore transcriptional activity hybrid promoters were constructed. The progressively deleted fragments of the hMTA-II_A regulatory region were fused to a 'crippled' HSV-TK promoter from the mutant pTKΔ5'-46 generated by McKnight¹³. The new hybrid promoters, therefore, contain a proximal promoter element from a gene (HSV-TK) not normally responsive to heavy metal ions¹⁰ and various portions of the 5' regulatory region of the hMT-II_A gene (Fig. 2a). The 3' deletion mutants are referred to as phMK-n1/-46, whereas n1 stand for the deletion end point within the hMT-II_A gene. phMK-n1/-46 series plasmids (phMK-n1/-46) were introduced into Rat 2 cells. Mixed cultures were derived from many TK⁺ colonies and analysed for expression of the chimaeric genes.

Surprisingly, we found that the 3' limit of the metal responsive element lies somewhere between positions -126 to -166, since phMK-126/-46 is Cd²⁺ inducible and phMK-166/-46 is not. The pTKΔ5'-46 promoter itself is not responsive to Cd²⁺ and its basal level of expression is lower than any one of the hybrid promoters (Fig. 2b). Similar results were obtained when the 3' deletion mutants were joined to the intact HSV-TK promoter from pTKΔ5'-109 (ref. 13), however, in this case, due to higher basal levels of expression, the fold induction after Cd²⁺ treatment was not as high (A.H., unpublished results and also see ref. 10).

In one series of experiments (Fig. 1b) the 5' limit of the metal responsive element seems to lie either at or 3' to position -50 and in another series of experiments (Fig. 2b) the 3' limit mapped between -126 to -166. However, this ambiguity was settled by comparison of the nucleotide sequences between positions -30 to -50 with the sequences between -135 to -155. An extensive homology of at least 12 nucleotides was found (Fig. 3). The same nucleotide sequence is shared with three other MT genes: mouse (m) MT-I (ref. 14), rat (r) MT-I (R. Andersen, personal communication) and hMT-I_A (ref. 15).

Close examination of the level of the 1.3 kilobase (kb) transcript initiated within the proximal HSV-TK promoter (data not shown but see ref. 10) relative to the level of the 0.9-kb transcript, which is initiated from a weak promoter within the structural portion of the HSV-TK gene (ref. 16), indicated that the basal level of expression of the hybrid promoter drops between phMK-43/-66 to phMK-43/-88. Therefore, the 3' limit of the element involved in maintaining the high basal level of expression of the hMT-II_A promoter or hybrid promoters is in between positions -66 to -88, the same region as defined earlier by the 5' deletions.

Binding site for hormone-receptor complex

To determine whether hormonal induction of the hMTK gene is a result of binding of the hormone-receptor complex to a specific site in the gene and whether the loss of biological response to glucocorticoids is correlated with loss of receptor binding to DNA we analysed the binding of partially purified glucocorticoid receptor from rat liver¹⁷ to the different deletion mutants. We found preferential binding of the receptor to those DNA fragments containing at least 268 bp of the region upstream of the hMT-II_A initiation site, whereas no selective binding was found to DNA fragments containing only 237 bp or less of this region (Fig. 4a and b). These findings suggest that the glucocorticoid responsive element between nucleotides -237 and -268 is also responsible for receptor binding to DNA. To define the glucocorticoid responsive sequence directly we performed DNase I protection experiments with the partially purified glucocorticoid receptor and end-labelled DNA fragments containing the hypothetical site^{18,19}. This showed a

Fig. 3 Consensus sequence of the heavy metal ion responsive element (MRE). Shown are the sequences between positions -51 to -35 (hMT-II_A no. 1) and -151 to -136 (hMT-II_A no. 2) of the hMT-II_A gene⁹ and the relevant sequences on the hMT-1_A (ref. 15) and the mMT-1 (ref. 14) genes. Underneath is shown the consensus sequence derived from comparison of those four regions.

hMT-II _A 1	ACTCGTCCCGG.CTCTTT
	-50 -40
hMT-II _A 2	GTGCG.CCCGG.CCCAGT
	-150 -140
hMT-1 _A	TTGCG.TCCGG.CCCTCT
	-50
mMT-1	TTGCG.CCCGGACTCGTC
	-50 -40
CONSENSUS	TG T
	--CG.CCCGG.C-C
	CT C

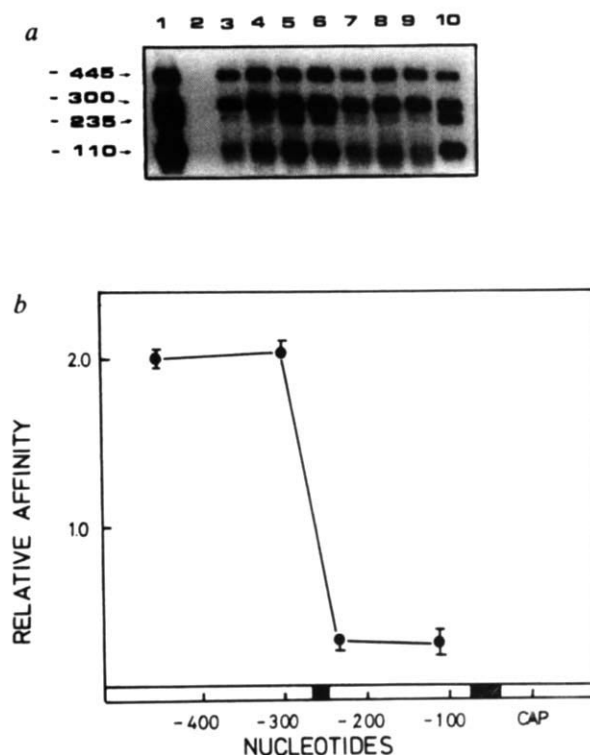
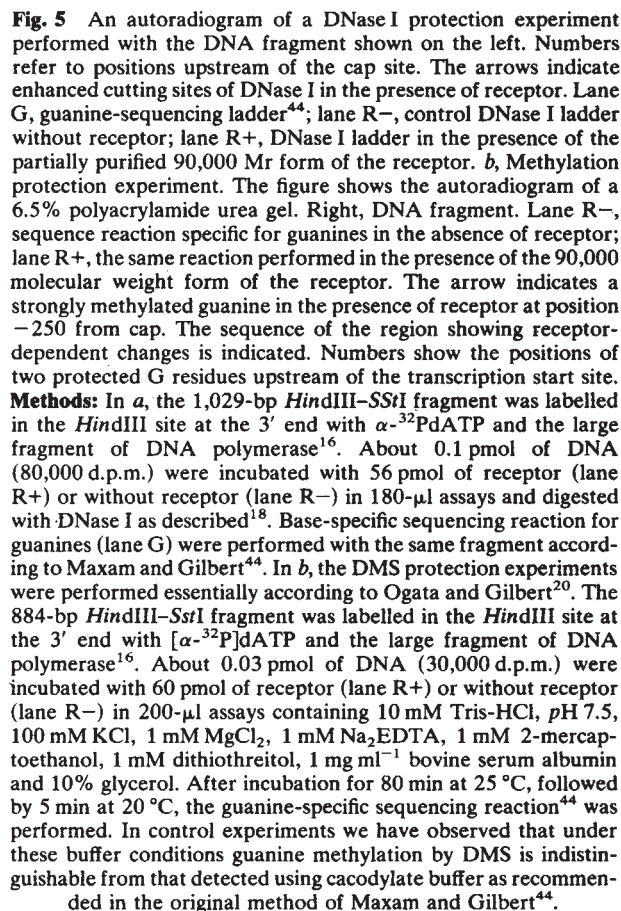


Fig. 4 Filter-binding experiments with hMT-II_A-TK deletion mutants. Four chimaeric plasmids in which the hMT-II_A-insert had been deleted up to the indicated positions upstream of the cap site (Fig. 1) were digested with *Hind*III and *Sst*I which cuts within the HSV-TK gene. The fragments were isolated on low-melt agarose gels, mixed in equimolar amounts and 3' end-labelled at the *Hind*III site with [α -³²P]dATP and the large fragment of *Escherichia coli* DNA polymerase¹⁶. The whole length of the four fragments were 1,029 bp (deleted to -445), 884 bp (deleted to -300), 819 bp (deleted to -235) and 694 bp (deleted to -110) respectively. **a**, An autoradiogram of the DNA fragments bound to nitrocellulose filters by the partially purified 90,000 molecular weight form of the receptor and separated by agarose gel electrophoresis. 5 ng DNA were used for each assay. Lanes 1 and 10 show the input DNA mixture. Lane 2 shows the DNA retained on the filter in the absence of receptor. Lanes 3-5 show the fragments retained at 60 mM KCl with increasing receptor concentrations (lane 3: 12 ng; lane 4: 28 ng; lane 5: 43 ng of receptor). Lanes 6 and 7 show the DNA fragments retained after incubation with 43 ng receptor at 90 mM KCl (lane 6) or 120 mM KCl (lane 7). Lanes 8 and 9 show the DNA bound to the filter by 43 ng receptor at 60 mM KCl in the presence of 25 ng native calf thymus DNA (lane 8) or 100 ng native calf thymus DNA (lane 9). **b**, Densitometric evaluation of the autoradiogram shown in **a**. The abscissa shows the region upstream of the transcription initiation site of hMT-II_A (cap). The numbers refer to the nucleotides retained in the corresponding deletions. The ordinate represents the relative affinity, defined as the ratio between the fractional intensities of the individual bands in lanes 3, 7, 8 and 9 of the autoradiogram and the fractional intensities of the same bands in the input DNA. Average and range are shown. The black box indicates the glucocorticoid regulatory element and the cross-striped box the region required for Cd regulation.



Taken together, these results demonstrate unequivocally that the same DNA region (-265 to -245) acts both as a regulatory element necessary for glucocorticoid induction of transcription, defined by a reversed genetic approach, and as a binding site for the hormone-receptor complex, defined by the DNA binding studies. Therefore the *in vitro* determined binding site is probably identical to the physiological site of interaction between the receptor and the gene. Moreover, since the deletion of that site does not effect the basal level of pHMT-II_A promoter activity or its ability to be induced by heavy metal ions, it appears that the glucocorticoid hormone-receptor is a positive regulatory factor. This is the first time such a relationship has been clearly established. Similar results, have recently been obtained by several groups investigating both hormone-receptor binding and expression of the long terminal repeat (LTR) of mouse mammary tumour virus (MMTV)^{17,19,21-23}. However, due to certain inherent limitations of the MMTV system, the relationship between receptor binding sites determined *in vitro*, and effects of hormone on transcription determined *in vivo* is not as clear as for the hMT-II_A promoter. Due to very low level of basal promoter activity in the absence of hormone, it is hard to dissociate the effects of deletions on inducibility of the viral

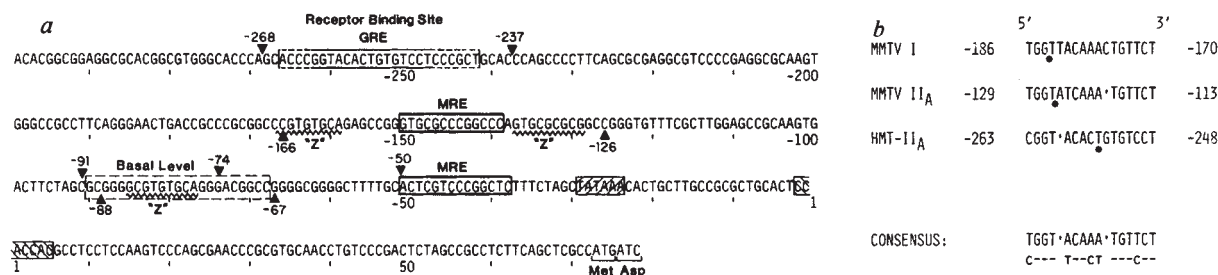


Fig. 6 *a*, Molecular anatomy of the *hMT-II_A* promoter. Sequence of the sense strand of DNA between positions -300 and +80. Regions important for hormonal induction (GRE for glucocorticoid responsive element), heavy metal induction (MRE for metal ion responsive element), and basal level of expression. Glucocorticoid regulatory element: solid line, the region protected by binding of the receptor to DNA; broken line, uncertainties in exact borders of the region. Also indicated are the TATA box and the cap site (hatched boxes) and region of alternating purine-pyrimidine sequences (Z-DNA), and the first two amino acid codons of the translated region. Arrowheads and numbers indicate the positions of deletion end points important for the mapping of the various regulatory elements. Arrowheads above the line indicate 5' deletions, and below the line 3' deletions. *b*, Consensus sequence for the glucocorticoid receptor binding to DNA. The sequence designated MMTVI and MMTVII_A correspond to regions in the LTR of MMTV strongly protected by the receptor against DNase I digestion¹⁹. The *hMT-II_A* sequence is taken from Fig. 4a and ref. 9. The asterisks mark the symmetry centre of small palindromes.

promoter, from an effect on its basal transcriptional activity^{21,23}. Also, the small, but significant induction observed in an MMTV deletion mutant containing only 50 bp of 5' flanking DNA²¹ and presumably no receptor binding sites^{18,19} leads to further confusion.

Comparison of the *hMT-II_A* gene nucleotide sequence at the region between -267 and -243 (ref. 9) to the MMTV-LTR sequences between -191 to -170 and -132 to -113 that are strongly protected against DNase I digestion¹⁹ reveals a high degree of nucleotide sequence homology. We propose that the conserved sequence elements between those three different binding sites (shown in Fig. 6b) make up the consensus nucleotide sequence for glucocorticoid hormone-receptor binding to DNA and consequently hormonal induction. Final confirmation, however, will depend on the outcome of experiments testing the biological activity of synthetic consensus sequences. Interestingly, a sequence which is in complete agreement with our proposed consensus sequence is found within the first intron (at position +95 to +109) of the human growth hormone gene²⁴, a gene which is induced by dexamethasone after transfection²⁵. Preliminary results indicate that this sequence functions as a receptor binding site (D. Moore, personal communication).

A weaker binding of the receptor to DNA is observed when the hexanucleotide 5'-TGT^TCT-3' occurs isolated from the rest

of the consensus sequence, as for instance between -70 and -100 in MMTV-LTR or at -324 in the antisense strand of *hMT-II_A* (Figs 4a and 5a). Since, however, this region of the *hMT-II_A* gene can be deleted without influencing the extent of hormonal induction, it seems that at least in this case binding of the receptor to the hexanucleotide is not functionally relevant.

Also we note that binding of receptor to the hexanucleotide results in a protected region covering only 12 bp, while binding to the complete consensus sequence protects 24 bp (Fig. 5a and ref. 19). It is likely that a dimer of the receptor is involved in generating a biological response. Methylation protection experiments suggest that at least two G residues located at -249 and -258 participate in the interaction of the receptor with its binding site. Since these residues are separated by 9 bp they should be located on the same side of the DNA helix. In addition, DMS methylation is selective for the N-7 position of G residues located in the DNA major groove²⁶. Thus, our results suggest that the receptor contacts at least one face of the DNA helix via the major groove.

We believe that a third control element is located between nucleotides -90 and -70, since 5' and 3' deletion mutants lacking this region exhibit reduced basal level of expression. A requirement for certain upstream nucleotide sequences, not directly involved in heavy metal induction, but which are important for

maintenance of basal level of expression was seen by Palmiter and Hamer and co-workers (ref. 27. P. Searle; A. Carter and D. Hamer, personal communications) studying the mMT-I gene by various transient expression assays.

Interestingly, the region between nucleotides -70 to -90 contains a stretch of alternating purine-pyrimidine residues: GCGTGTGCA and a similar sequence is located 80-bp upstream: CGTGTGCA. Such sequences capable of forming a Z-DNA structure have been found in a large number of transcriptional enhancer sequences from a variety of DNA and RNA tumour viruses²⁸. A functional role for these elements is strongly suggested by the analysis of the transcriptional activity of the deletion mutants. While deletion of the upstream segment does not have a significant effect on the basal level of gene expression, deletion of both elements dramatically reduces the basal activity of the *hMT-II_A* promoter.

The overall organization of the *hMT-II_A* gene regulatory region resembles that of several other eukaryotic genes²⁹⁻³⁴. In addition to the TATA box region, a functional element common to most eukaryotic genes³⁵, those genes were found to have at least two other functional elements. One of them is usually located near the TATA box region, and the other located further upstream is in the -90 region. In addition, further upstream elements required for efficient transcription have been described for the sea urchin H2A gene³⁶ and the simian virus 40 early region^{32,34,37}.

As several inducible higher eukaryotic genes have now been analysed by *in vitro* mutagenesis and gene transfer, some conclusions concerning control of transcription can be drawn. First, unlike prokaryotic promoters, the eukaryotic promoter is extended over a large region composed of different functional elements (270 bp in the case of the *hMT-II_A* gene). Second, the comparison of functionally related control elements reveals the conservation of nucleotide sequences among such elements (refs 38-40 and this study). The conservation of nucleotide sequences among functionally related regulatory elements suggests that those elements serve as binding sites for various regulatory proteins. The identity of most regulatory proteins involved in the various induction phenomena mentioned above, is not known yet. However, our current study and recent work on MMTV-LTR^{19,21-23} have established that one such protein, the glucocorticoid receptor, interacts directly with regulatory sites important for the biological response.

How is this interaction translated into increased transcription (or repression in some cases)? The exact molecular details are not clear yet, but recent work on the regulatory elements of *hMT-II_A* gene¹⁰, the MMTV-LTR (refs 21-23) and of the *Hsp70* heat shock gene^{39,40} indicates that those elements can activate heterologous promoters when placed at various distances upstream. This feature bears some resemblance to the

effects of a better known class of regulatory elements, the viral 'enhancers'^{32,34,37,41}. It is quite possible that the various 'modulator' elements, described, can serve as regulated RNA polymerase entry sites. Another possibility, which we favour, is that the interaction of regulatory proteins with the DNA at upstream sites leads to a change in the conformation of the double helix, a signal that can be transmitted over some distance, and lead

to local melting at the TATA box region, the site for RNA polymerase and initiation factors binding to DNA⁴².

We thank Uwe Vaupel for technical assistance, Claus Scheidereit and Dr Manuel Grez for helpful suggestions. This work was supported by grants from the Deutsche Forschungsgemeinschaft, the EPA (R-810531-01) and the NIH (R01-ES-03222-01).

Received 26 September 1983; accepted 6 January 1984.

1. Kagi, J. H. R. & Nordberg, M. (eds) *Metallothionein* (Birkhauser, Basel, 1979).
2. Karin, M. & Herschman, H. R. *Science* **204**, 176-177 (1979).
3. Mayo, K. E. & Palmiter, R. D. *J. biol. Chem.* **256**, 2621-2624 (1981).
4. Durnam, D. M. & Palmiter, R. D. *J. biol. Chem.* **256**, 5712-5716 (1981).
5. Karin, M. *et al. Nature* **286**, 295-297 (1980).
6. Hager, L. J. & Palmiter, R. D. *Nature* **291**, 340-342 (1981).
7. Karin, M., Slater, E. P. & Herschman, H. R. *J. cell. Physiol.* **106**, 63-74 (1981).
8. Karin, M., Andersen, R. D. & Herschman, H. R. *Eur. J. Biochem.* **118**, 527-531 (1981).
9. Karin, M. & Richards, R. *Nature* **299**, 797-802 (1982).
10. Karin, M. *et al. Cell* **36**, 371-379 (1984).
11. Gray, H. B. *et al. Nucleic Acids Res.* **2**, 1459-1492 (1975).
12. Weaver, R. F. & Weissman, C. *Nucleic Acids Res.* **5**, 1175-1193 (1979).
13. McKnight, S. L., Gavis, E. R., Kingsbury, R. & Axel, R. *Cell* **25**, 385-398 (1981).
14. Glanville, N., Durnam, D. M. & Palmiter, R. D. *Nature* **292**, 267-269 (1981).
15. Richards, R. L., Heguy, A. & Karin, M. *Cell* (submitted).
16. Roberts, J. M. & Axel, R. *Cell* **29**, 109-119 (1982).
17. Geisse, S. *et al. EMBO J.* **1**, 1613-1617 (1982).
18. Galas, D. & Schmitz, A. *Nucleic Acids Res.* **5**, 3157-3170 (1978).
19. Scheidereit, C. *et al. Nature* **304**, 749-752 (1983).
20. Ogata, R. G. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5851-5854 (1978).
21. Hynes, N. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 3637-3641 (1983).
22. Chandler, V., *et al. Cell* **33**, 489-499 (1983).
23. Majors, J. & Varmus, H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5866-5870 (1983).
24. DeNoto, F. M., Moore, D. D. & Goodman, H. M. *Nucleic Acids Res.* **9**, 3719-3730 (1981).
25. Robbins, D. M., Paek, I., Seeburg, P. H. & Axel, R. *Cell* **29**, 623-631 (1982).
26. Siebenlist, U., Simpson, R. & Gilbert, W. *Cell* **20**, 269-281 (1980).
27. Brinster, R. L. *et al. Nature* **269**, 39-42 (1982).
28. Nordheim, A. & Rich, A. *Nature* **303**, 674-679 (1983).
29. McKnight, S. L. & Kingsbury, R. *Science* **217**, 316-324.
30. Dierks, P. *et al. Cell* **32**, 695-706 (1983).
31. Grosfeld, G. C., de Boer, E., Shewmaker, C. K. & Flavell, R. A. *Nature* **295**, 120-126 (1982).
32. Benoist, C. & Chambon, P. *Nature* **290**, 304-310 (1981).
33. Everett, R. D., Baty, D. & Chambon, P. *Nucleic Acids Res.* **11**, 2447-2464 (1983).
34. Fromm, M. & Berg, P. *J. molec. appl. Genet.* **1**, 457-481 (1982).
35. Goldberg, D. thesis, Stanford Univ. (1980).
36. Grosschedl, R. & Birnstiel, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1432-1438 (1980).
37. Gruss, P., Dhar, R. & Khoury, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 943-947 (1981).
38. Ragg, H. & Weissmann, C. *Nature* **303**, 439-442 (1983).
39. Pelham, H. R. B. *Cell* **30**, 517-528 (1982).
40. Pelham, H. R. B. & Bionz, M. *EMBO J.* **1**, 1473-1477 (1982).
41. Banerji, J., Rusconi, S. & Shaffner, W. *Cell* **27**, 299-308 (1981).
42. Davison, B. L., Egly, J. M., Mulvihill, E. R. & Chambon, P. *Nature* **301**, 680-686 (1983).
43. Vieira, J. & Messing, J. *Gene* **19**, 259-268 (1982).
44. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499-559 (1980).
45. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).

LETTERS TO NATURE

X-ray emission from non-magnetic cataclysmic variables

A. R. King & G. Shaviv*

Astronomy Department, University of Leicester,
Leicester LE1 7RH, UK

Many (~50) cataclysmic variables, binary systems in which a white dwarf accretes from a close companion star, are now known to be hard (≥ 2 keV) and variable X-ray sources¹. While some of these (~15) exhibit regular pulsing in X rays, suggesting that they have magnetic fields strong enough to channel the accretion flow onto restricted parts of the white dwarf surface (the AM Her and 'intermediate polar' systems), the majority do not, and hence presumably have rather weaker fields ($B \leq 10^5$ G). SS Cygni, the brightest object of this 'non-magnetic' class, exhibits a combination of systematic features in its X-ray and optical emission which existing models are only partially successful in reproducing. We suggest here that at sufficiently low accretion rates a thermal instability in the accretion flow leads to the formation of a hard X-ray-emitting corona around the white dwarf; the corona is relatively suppressed at higher accretion rates. We show that the cooling of such a corona gives the observed correlation between X-ray temperature and luminosity, and that hard X-ray production may be rather inefficient because of transport processes. These can give rise to soft X-ray emission from most of the white dwarf surface, with a correspondingly low effective temperature.

Any theoretical model for the hard X-ray emission from these systems must explain three key observational facts: (1) SS Cygni is a bright and very variable hard X-ray source in quiescence (that is, low accretion rates $\dot{M}$), but the hard X rays are severely reduced during optical outbursts² (increased accretion rates) while the soft X rays brighten^{1,2}. (2) The drop in hard X-ray intensity of SS Cygni is not accompanied by an increase in intrinsic (for example, photoelectric) absorption³. (3) Hard X-ray spectra of SS Cygni are thermal, with bremsstrahlung tem-

perature $kT \leq 20$ keV¹; these are strongly correlated (J. H. Swank, cited in ref. 1 and M. J. Ricketts and A. Bennetts, personal communication) with X-ray luminosity L_x , T and L_x rising and falling together.

The only plausible energy reservoir for hard X-ray production is the boundary layer where the inner edge of the accretion disk interacts with the white dwarf. If this has mass and radius M , R_* , the keplerian kinetic energy which must be lost by the boundary-layer material before accreting to the white dwarf (assumed to be rotating below break-up velocity) provides a luminosity $GMM/2R_*$. Existing models^{4,5} of hard X-ray production invoke optically thin emission from the boundary layer; if the boundary layer is optically thick, soft X rays at an effective temperature $kT \sim 5$ –10 eV result⁶. Refs 4, 5 both explain fact (1) above in terms of increased absorption, and hence fall foul of fact (2). While Tytenda's model⁵ gives no natural explanation for fact (3), we shall show that this is likely to follow for any model in which the X-ray-emitting gas can expand to form a corona around the white dwarf before cooling. This is the case in Pringle and Savonije's model⁴ for low $\dot{M}$.

We consider an optically thin element of boundary-layer gas of radial extent $b < R_*$, orbiting the white dwarf with azimuthal velocity $v_\phi \sim v_k = (GM/R_*)^{1/2}$ (by definition $v_\phi < v_k$ in the boundary layer). Because of its high rate ($\sim v_\phi/b$) of shear, we expect the boundary-layer gas to be very turbulent; this turbulence will both decelerate and heat the orbiting gas. We write

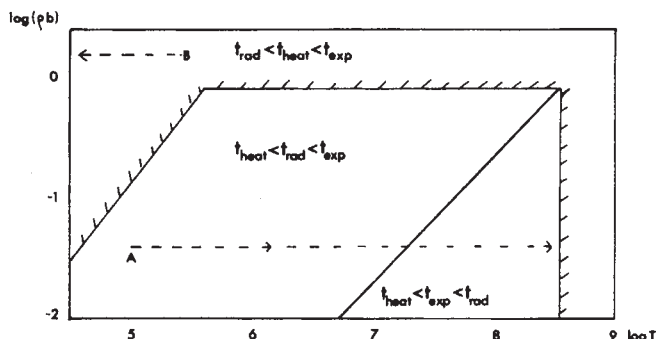


Fig. 1 Heating plane for the case $M = M_\odot$, $R = 5 \times 10^8$ cm, $\alpha = 1$, $(v_\phi/v_k)^2 = 0.5$. Gas elements characterized by initial points lying in the hatched region heat up to X-ray temperatures.

* Permanent address: Faculty of Physics, Technion, Haifa, Israel.

the local volume heating rate Γ ($\text{erg s}^{-1} \text{cm}^{-3}$) in the form

$$\Gamma = \nu \rho v_\phi^2 / b^2 \quad (1)$$

with ρ the density and ν a kinematic eddy viscosity. As usual, we can parametrize ν as

$$\nu = \alpha c_s b \quad (2)$$

where c_s is the local sound speed and α a dimensionless parameter, of order unity, for sonic turbulence. If no other processes operated, the gas would heat up on a time scale

$$t_{\text{heat}} = 3\rho c_s^2 / \Gamma \quad (3)$$

However, the gas may expand on a time scale

$$t_{\text{exp}} = b / c_s \quad (4)$$

and cool radiatively on a time scale

$$t_{\text{rad}} = 3\rho c_s^2 / \rho^2 \Lambda \quad (5)$$

where $\rho^2 \Lambda(T)$ ($\text{erg s}^{-1} \text{cm}^{-3}$) is the volume loss rate. From equations (1)–(4) we find

$$\frac{t_{\text{heat}}}{t_{\text{exp}}} = \frac{3c_s^2}{\alpha v_\phi^2} \quad (6)$$

Using $c_s^2 = \gamma kT / \mu m_H$ this shows that adiabatic expansion prevents the gas being heated above the 'virial' temperature

$$T_v = \alpha \mu m_H v_\phi^2 / 3\gamma k \quad (7)$$

For $\alpha \sim 1$, T_v is, of course, of the order of the temperature T_s ($\sim 10^8 \text{ K}$) achieved by strongly shocking the azimuthal velocity v . From equations (1)–(5) we find similarly

$$\frac{t_{\text{rad}}}{t_{\text{heat}}} = \frac{\alpha c_s v_\phi^2}{\rho b \Lambda} \quad (8)$$

and

$$\frac{t_{\text{rad}}}{t_{\text{exp}}} = \frac{3c_s^3}{\rho b \Lambda} \quad (9)$$

Since c_s and Λ are functions of T only, the ratios (6), (8) and (9) depend only on the variables T and ρb , and we may plot a 'heating plane' (Fig. 1). Here we have used $\Lambda(T)$ approximated as a pair of power laws in T , from Fig. 3 of Buff and McCray⁷, as appropriate for the highly ionized boundary-layer gas.

Let us consider the fate of an element of boundary-layer gas using Fig. 1. We expect the element to start from a temperature $T \sim 10^5 \text{ K}$ as this is the temperature of nearby optically thick matter in the disk, boundary layer and white dwarf envelope. If ρb is sufficiently low, the element will correspond initially to a point such as A inside the hatched region. Here $t_{\text{heat}} < t_{\text{rad}} < t_{\text{exp}}$, so the element heats up and moves to the right in the figure, until it crosses the line $t_{\text{rad}} = t_{\text{exp}}$ given by equation (9). In this region $t_{\text{heat}} < t_{\text{exp}} < t_{\text{rad}}$, so the element continues to heat up, but now tries to expand rather than radiate. Finally it reaches the vertical line $T = T_v$ on which $t_{\text{heat}} = t_{\text{exp}}$ [equation (7)]. Here further heating becomes ineffective and the element expands with $T \approx T_v$. Gas at such temperatures has a 'vertical' or 'disk' scale height $\sim (c_s/v_k) R_* \sim 0.6 R_*$, so it cannot be confined to the disk plane and must form an X-ray-emitting corona around the whole white dwarf, of radial scale height $H_r \sim (c_s/v_k)^2 R_* \leq 0.3 R_*$. We shall investigate the cooling of such a corona below. For elements with larger initial values of ρb , such as that represented by point B in Fig. 1, $t_{\text{rad}} < t_{\text{heat}} < t_{\text{exp}}$, and the gas can cool radiatively as fast as it is heated. Such elements collapse onto the white dwarf surface without ever reaching hard X-ray temperatures.

It is not difficult to see here a potential explanation of the observational features (1) and (2) above: at low accretion rates, most of the boundary-layer gas is presumably characterized by small values of ρb and must therefore feed a hard X-ray corona. As $\dot{M}$ rises, the extra mass of gas in the boundary layer will at first lead to an increase in X-ray intensity, but eventually more

and more of the gas will probably have values of ρb placing it outside the hatched region of the heating plane (Fig. 1). If $\dot{M}$ gets high enough, very little gas will lie in the hatched region, and hard X-ray emission will be suppressed. Note that this also agrees with point (2) above, in that no absorption is required to blot out hard X rays; they are simply not produced in great quantity for high accretion rates. Note also that any weak shocks in the boundary layer which heat the gas above 10^5 K will increase the likelihood of its initial position lying in the hatched region of the heating plane. Also, the weakly-bound coronal gas offers a convenient site from which to blow away a wind, as sometimes observed in cataclysmic variables (see, for example, ref. 1).

Let us consider now the X-ray coronal gas. For this, two relevant time scales are the radiative cooling time t_{rad} to bremsstrahlung X-ray emission and the hydrodynamic time $t_{\text{hyd}} = H_r / c_s = c_s R_* / v_k^2$. In the analogous case of the solar corona a further important time scale is that of conductive cooling (for example, ref. 8)

$$t_{\text{cond}} = \frac{5\rho c_s^2 H_r^2}{2\kappa_0 T^{7/2}} \quad (10)$$

where $\kappa_0 = 10^{-6}$ (CGS units). Clearly, conduction can only be important in non-steady-state situations: in a steady state there will be a temperature minimum between the corona and the white dwarf photosphere, across which nothing can be transported by diffusive thermal conduction⁹. (This does not apply to energy transport by high-velocity shock-heated electrons: refs 9–11 show how 'accretion column' solutions can be constructed using this process.) Our white dwarf corona, like that of the Sun, will be highly unsteady, and can be thought of as a superposition of flares; this agrees with the observed strong variability of hard X rays from, for example, SS Cygni. The main uncertainty as to whether conduction is important arises because we do not know the surface magnetic field configuration of the white dwarf. Although the field is assumed to be dynamically unimportant for the accretion flow, it may still be capable of suppressing transport processes, especially conduction, across fieldlines. In the Sun this means that conduction acts one-dimensionally along coronal loops, and only heats the ends of these loops. The same could be true in our case, or it is conceivable that the field somehow prevents conduction from working effectively at all. We shall consider both possibilities in what follows; our explanation of point (3) is the same in either case, but conduction can have a marked effect on the total X-ray luminosity, as we shall see.

We construct the ratios $t_{\text{rad}}/t_{\text{hyd}}$, $t_{\text{rad}}/t_{\text{cond}}$, $t_{\text{hyd}}/t_{\text{cond}}$ and plot a 'cooling plane' (Fig. 2) (note that the variables here are T and ρ , rather than T and (ρb) as for the heating plane, Fig. 1). We expect coronal gas elements to start with temperatures $\approx T_v$. For low enough initial densities ρ ($\leq 5 \times 10^{-10} \text{ g cm}^{-3}$ in Fig. 2), the gas elements start at points like A. Here $t_{\text{cond}} < t_{\text{hyd}} < t_{\text{rad}}$, so that T decreases, while ρ and H_r , which vary on time scales $\sim t_{\text{hyd}}$, are left unchanged. The gas element therefore cools to point B by conduction with $L_x \propto H_r \rho^2 T^{1/2} \sim T^{1/2}$: the lines $L_x(\text{max}) = 4\pi R^2 H_r \rho^2 \Lambda$ on Fig. 2 show that this is likely to reduce very sharply the luminosity available to be radiated as hard X rays. (The heat conducted into the white dwarf will mostly be radiated from its surface as black-body-like radiation of effective temperature $\sim 10^5 \text{ K}$, see below.) At points to the left of B we have $t_{\text{hyd}} < t_{\text{cond}} < t_{\text{rad}}$: here the density and scale-height have time to respond to the loss of energy by conduction. This is the regime investigated extensively in the study of solar flares⁸. The 'thermal overload' due to the conductive heating is likely to lead to evaporation of matter at the base of the corona: the result is that as T decreases further, H_r decreases ($\sim T$) and ρ increases somewhat. Thus $L_x \propto H_r \rho^2 T^{1/2} \sim T^\gamma$, with γ varying from $-\frac{1}{2}$ to $+\frac{3}{2}$. Finally, the temperature reaches point C; here $t_{\text{hyd}} < t_{\text{rad}} < t_{\text{cond}}$, and we have a hydrostatic gas cooling radiatively. Hence H_r continues to decrease like T , but now ρ also decreases slowly with T , giving $L_x \propto T^{3/2+2\epsilon}$, $\epsilon > 0$. To under-

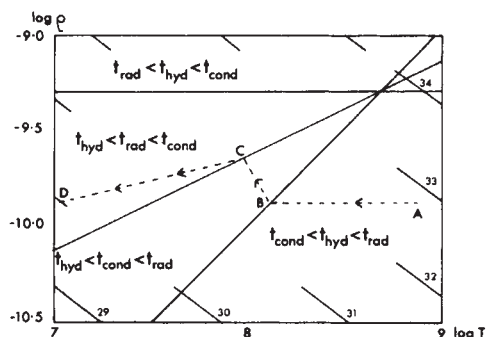


Fig. 2 Cooling plane for the case $M = M_{\odot}$, $R = 5 \times 10^8$ cm. Low-density coronal gas follows tracks like A-B-C-D, spending most of its time in region C-D, where $L_x \propto T^{3/2+2\epsilon}$. Also shown are lines $L_x = \text{constant}$ (broken, for clarity), labelled by the value of $\log L_x$.

stand the density behaviour, we note that the hydrostatic and energy equations are

$$\frac{\partial P}{\partial z} = -g\rho$$

$$\frac{3}{2} \frac{\partial P}{\partial t} = -\rho^2 \Lambda$$

with $g = GM/R_*^2$, P the gas pressure and z the vertical coordinate. Thus

$$\frac{\partial \rho}{\partial t} = -\frac{1}{g} \frac{\partial^2 P}{\partial t \partial z} = \frac{2}{3g} \frac{\partial}{\partial z} (\rho^2 \Lambda) < 0$$

since ρ must decrease with z , while T is effectively constant with z . This situation will persist, with faster cooling at the base of the corona causing matter to condense there and reducing ρ in the X-ray-emitting corona above. Thus, low-density coronal gas will pass successively through a conductive cooling phase with $L_x \propto T^{1/2}$, a 'solar flare' phase, with $L_x \propto T^\gamma$, $-\frac{1}{2} \leq \gamma \leq \frac{3}{2}$, and a hydrostatic cooling phase, with $L_x \propto T^{3/2+2\epsilon}$. The last phase is the longest, and is also much longer than the heating phase. Therefore, with on average a roughly constant mass in the corona, most observed X-ray emission will come from gas in this phase. Thus, observationally we expect $L_x \propto T^{3/2+2\epsilon}$, $\epsilon > 0$, in agreement with fact (3) above. The actual value of the temperature exponent, $\frac{3}{2} + 2\epsilon$, and the range of T and L_x , is encouragingly close to what is seen for SS Cygni (M. J. Ricketts and A. Bennetts, personal communication). If conduction is ineffective, this low-density gas would spend all its time in regions like C-D, with $L_x \propto T^{3/2+2\epsilon}$. Effective conduction may be desirable, however, as it offers an explanation of the rather low (10^{30} – 10^{32} erg s $^{-1}$) hard X-ray luminosities (see Fig. 2) and the fact that a soft X-ray component is unobservable in most sources. This follows, since if much of the boundary-layer energy release is stored in the corona and then transported by conduction into the white dwarf, the area emitting this luminosity in black-body form is a large fraction of the white dwarf surface, rather than an area comparable with that covered by the boundary layer. The resulting effective temperature $T_b \sim 3 \times 10^4$ K is too soft for detection with existing satellites ($kT_b \sim 3$ eV) except for some bright (or outbursting) systems.

The lack of a detailed model of the boundary-layer flow means that we cannot predict how much of the coronal gas will be at the required low densities ($\leq 5 \times 10^{-10}$ g cm $^{-3}$ in Fig. 2). At higher densities t_{rad} is always the shortest time scale ($T < 10^9$ K) and the gas cools with constant ρ and H_r , giving $L_x \propto T^{1/2}$. Any such high-density gas would, of course, cool faster and hence be less noticeable in the integrated emission. It will be interesting to see, however, if any systems are observed to show the weaker ($L_x \propto T^{1/2}$) dependence predicted if most of the coronal gas has a high density (it may also be possible to see changes in the temperature dependence of L_x with X-ray brightness in a given

source). Quite possibly the accretion rates required to provide such densities would suppress the heating instability which gives rise to the X rays in the first place. Expressing the average mass in the corona in two ways (ref. 4) we find $\rho \sim 10^{-10} M_{16}^{1/2}$ g cm $^{-3}$ for typical parameters, with $M_{16} = M/10^{16}$ g s $^{-1}$. A theoretical priority must be the construction of a model for the boundary-layer flow. Since, however, the accretion process is clearly very time-dependent, and definitely not in a steady state, this is no simple task.

G.S. thanks the SERC for Visiting Fellowship support. We thank A. Bennetts and M. J. Ricketts for discussion of the X-ray data on SS Cygni. A.R.K. thanks members of the astrophysics groups at the University of Bonn and MPI, Munich, for discussion at an early stage of this work.

Received 10 November 1983; accepted 1 February 1984.

1. Córdova, F. A. & Mason, K. O. in *Accretion Driven Stellar X-ray Sources* (eds Lewin, W. H. G. & van den Heuvel, E. P. J.) Ch. 4 (Cambridge University Press, 1983).
2. Ricketts, M. J., King, A. R. & Raine, D. J. *Mon. Not. R. astr. Soc.* **186**, 233 (1979).
3. Swank, J. H. *IAU Colloq.* **53**, 135 (1979).
4. Pringle, J. E. & Savonije, G. J. *Mon. Not. R. astr. Soc.* **187**, 777 (1979).
5. Tytenda, R. *Acta astr.* **31**, 127 (1981).
6. Pringle, J. E. *Astr. Astrophys.* **19**, 137 (1981).
7. Buff, J. & McCray, R. *Astrophys. J.* **189**, 147 (1974).
8. Pallavicini, R. *et al. Astrophys. J.* **270**, 270 (1983).
9. Frank, J., King, A. R. & Lasota, J. P. *Mon. Not. R. astr. Soc.* **202**, 183 (1983).
10. King, A. R. *IAU Colloq.* **72**, 181 (1983).
11. Frank, J. & King, A. R. *Astr. Astrophys.* (in the press).

A simple correlation between isotropic ^{29}Si -NMR chemical shifts and T-O-T angles in zeolite frameworks

S. Ramdas & J. Klinowski

Department of Physical Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EP, UK

^{29}Si high-resolution solid-state nuclear magnetic resonance (NMR) spectrometry with magic-angle spinning (MASNMR) has proved to be a valuable tool for the quantitative identification of local $\text{Si}(\text{O-Al})_n(\text{O-Si})_{4-n}$ environments of silicon in zeolitic frameworks $^{1-4}$ where $n = 0, 1, 2, 3$ or 4. In addition, in certain cases, notably silicalite 5 and several highly siliceous (dealuminated) zeolites 6 , crystallographically non-equivalent $\text{Si}(\text{O-Si})_4$ sites have also been distinguished. The accumulated NMR data on a number of zeolites have enabled us to derive a semi-empirical relationship between ^{29}Si isotropic chemical shifts in aluminosilicates and the non-bonded Si...T distance ($T = \text{Si}$ or Al). The relationship holds for all aluminosilicates and all the five types of $\text{Si}(\text{O-Al})_n(\text{O-Si})_{4-n}$ tetrahedral environments; it also applies to silica polymorphs. The average value of Si-O-T angle for each kind of silicon site in aluminosilicates of unknown structure can now be estimated from ^{29}Si -MASNMR, which holds considerable promise for structure elucidation.

The magnitude of ^{29}Si isotropic chemical shifts in silicates and aluminosilicates was at first thought to be structure-independent, but it was soon shown that shifts for the same type of tetrahedral environment in zeolites Na-A and Na-X differ by ~ 6 p.p.m., so that the $\text{Si}(4\text{Al})$ signal in the former material coincides with $\text{Si}(3\text{Al})$ in the latter. A similar situation was later found to obtain in other zeolites. To explain the unusual position of the $\text{Si}(4\text{Al})$ resonance in Na-A, we suggested 7 the existence of an effect of Si-O-T ($T = \text{Si}$ or Al) angle, θ , on the magnitude of the ^{29}Si chemical shift. Later studies by Jarman 8 on zeolites ZK-4 and TMA-sodalite, and by Grimmer *et al.* 9 on the mineral zunyite provide further support for this hypothesis. The linear ($\theta = 180^\circ$) Si-O-Si linkage in zunyite, causes the ^{29}Si resonance to occur at the unusually negative value of -128.5 p.p.m. from

Table 1 Chemical, structural and NMR spectroscopic parameters of various silicates and aluminosilicates

No.	Material*	Chemical shift (p.p.m. from TMS) for various Si(<i>n</i> Al) environments§					Si-O-T angle θ (deg)	Refs¶
		Si(4Al)	Si(3Al)	Si(2Al)	Si(1Al)	Si(0Al)		
1	Li-ABW (1.0)	-80.0	—	—	—	—	135.0	#
2	Na-A (1.0)	-88.9	—	—	—	—	148.0	12, [13]
3	Li-A (1.0)	-85.0	—	—	—	—	141.0	#
4	ZK-4 (1.14)†	-89.1	-93.9	—	—	—	148.0	14
5	ZK-4 (1.14)†	-89.1	-93.9	-99.5	-106.1	—	148.0	14
6	ZK-4 (1.56)†	-88.4	-93.3	-99.3	-105.2	-111.0	148.0	7
7	Na-X (1.18)	-83.9	-88.1	—	—	—	139.2	15, [16]
8	Na-Y (2.75)‡	—	-89.6	-95.0	-100.4	-105.8	144.8	15, [17]
9	Cancrinite (1.0)	-87.2	—	—	—	—	142.3	18, [19]
10	Gmelinite (2.4)	—	-91.7	-97.1	-107.7	—	143.3	20, [21]
11	Chabazite (2.3)	—	-94.0	-99.4	-104.8	-110.0	145.4	2, [22]
12	Scolecite (1.5)	—	-86.3	—	—	—	138.3	
			-89.1	—	—	—	139.3	
			—	-95.8	—	—	147.4	23, [24]
13	Natrolite (1.5)	—	-87.7	—	—	—	137.2	
		—	—	-95.4	—	—	142.4	1, [25]
14	Laumontite (2.0)	—	—	-92.4	—	—	138.0	23, [26]
15	Analcime (2.0)	—	-92.0	-96.3	-101.3	—	144.3	2, [27]
16	Coesite	—	—	—	—	-108.1	144.2	
						-113.9	152.8	10
17	Cristobalite	—	—	—	—	-108.5	146.4	10
18	Quartz	—	—	—	—	-107.1	143.6	10
19	Tridimite	—	—	—	—	-111.0	150.0	10
20	Silicalite	—	—	—	—	-109.2	150.7	5, [28]
						-116.3	160.6	
21	TMA-sodalite	—	—	—	—	-116.2	157.8	8, [29]
22	Omega (mazzite)	—	—	—	—	-106.0	140.8	6, [30]
						-114.4	151.9	
23	Offretite	—	—	—	—	-109.7	142.5	
						-115.2	151.3	
24	Mordenite	—	—	—	—	-112.2	150.4	6, [31]
						-113.1	152.3	
						-115.0	156.0	6, [32]

* The Si/Al ratio is given in brackets for aluminosilicates. Materials 16–19 are polymorphs of silica, and 22–24 chemically treated (dealuminated) zeolites.

† It is assumed that the mean Si-O-T angle is invariant with increasing Si/Al ratio.

‡ The single Si(0Al) signal measured³³ at -107.4 p.p.m. in dealuminated zeolite Y is also in good agreement³⁴ with the predicted $\theta = 143.5^\circ$.

§ Only well resolved peaks of significant intensity are tabulated. When more than one type of crystallographic site is present, assignment of resonances is often possible when θ and population factors are taken into account.

|| The mean angle for each type of non-equivalent crystallographic site.

¶ Key references to NMR data (without brackets) and crystal structure parameters (in square brackets).

R. H. Jarman, J. M. Newsam and A. J. Jacobson, unpublished data.

tetramethylsilane (TMS). Smith and Blackwell¹⁰ recently reported good quantitative correlations between chemical shifts and Si···Si distances as well as the mean secant of the Si-O-Si angle for several polymorphs of silica. We prefer the former of the two correlations since it is capable of incorporating the difference between Si-O and Al-O bond lengths, essential to characterize all five types of Si(*n*Al) environments. The magnitudes of the ²⁹Si chemical shifts and of the Si-O-T angles have a common origin: the electronic configuration of the linkage. Taking the accumulated data on ²⁹Si chemical shifts in various silicates and aluminosilicates as a starting point, we are now able to introduce a simple semi-empirical relationship between δ and the average value of the Si-O-T angle in all these structures. The relationship is valid for chemical shifts in all the five Si(4Al), Si(3Al), Si(2Al), Si(1Al) and Si(0Al) tetrahedral environments, and also holds for the Si(0Al) shifts in silica polymorphs listed by Smith and Blackwell¹⁰.

While T-O-T angles in zeolites can be determined relatively reliably by conventional crystallographic methods, the occupancy (Si or Al) of tetrahedral sites is often more difficult to establish accurately. We base our argument on the assumption of constant average Si-O and Al-O bond length of 1.62 and 1.75 Å respectively¹¹, and proceed to reconstruct the (non-bonded) T···T distance between the two tetrahedral atoms. If

θ is the T-O-T angle, the Si···Si distance is $3.24 \sin(\theta/2)$ Å and the Si···Al distance $3.37 \sin(\theta/2)$ Å. The latter value is obtained by treating the Si-O-Al triangle as isosceles with the uniform T-O distance of 1.685 Å, that is, the average between Si-O and Al-O bond lengths. The Si···Al distance thus obtained agrees to four significant figures with the value calculated rigorously by applying the cosine formula. The sum total of the four Si···T bond lengths involving the central Si atom in a Si(*n*Al) unit is therefore

$$\sum d_{TT} = [3.37n + 3.24(4 - n)] \sin(\theta/2) \quad (1)$$

Using the literature data on crystal structures and ²⁹Si-NMR chemical shifts, listed in Table 1, we plot δ as a function of $\sum d_{TT}$ for Si(0Al) units. Figure 1 shows that there is an approximately linear relationship and regression analysis of the data yields

$$\delta/\text{p.p.m.} = 143.03 - 20.34 \sum d_{TT}/\text{\AA} \quad (2)$$

with a correlation factor of 0.91. Assuming the same slope and including an additional term, $7.95n$, on the right-hand side of equation (2) to account for the paramagnetic contribution of *n* tetrahedral aluminium atoms to the chemical shift of the central silicon, we arrive at the remaining four lines in Fig. 1 for Si(*n*Al) units with *n* = 1, 2, 3 and 4. The corresponding experimental

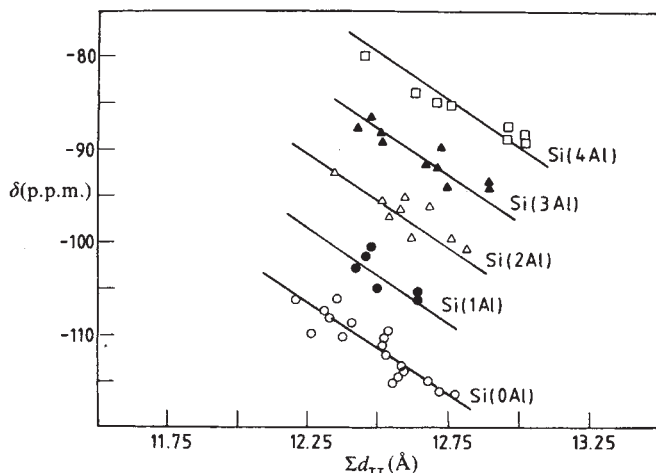


Fig. 1 The isotropic ^{29}Si -NMR chemical shift, δ (in p.p.m. from tetramethylsilane), plotted against Σd_{TT} (in Å), the calculated total Si···T non-bonded distance for Si atoms in the five kinds of Si(n Al) tetrahedral environments. Aluminosilicates corresponding to the various points can be identified by reference to Table 1.

data are close to the theoretical lines, despite the empirical character of the expressions.

Isotropic chemical shifts can be determined with an accuracy better than 1 p.p.m. while T-O-T angles are reported to about $\pm 2^\circ$. On the other hand, the scatter of points in Fig. 1 is sometimes greater than can be accounted for by these errors. Calculating Si···T distances from the average value of θ rather than from the separate Si-O-T angles θ_1 , θ_2 , θ_3 and θ_4 is mainly responsible for the error, especially in cases where the individual values of θ are very different. However, in most cases there is no alternative, as we do not know which angles correspond to Si-O-Si and which to Si-O-Al linkages. Second, the variation in T-O distances, which we assume to be constant, also leads to some small error. Typically, $d_{\text{Si-O}} = 1.62 \pm 0.01$ Å corresponds to an uncertainty in Σd_{TT} of ± 0.075 Å. Note that even a large uncertainty of 0.1 Å in Σd_{TT} corresponds to an error of only 0.9° in θ .

The above assumptions were made in order to keep equation (2) as simple as possible and thus to maximize its predictive power. We submit that on the basis of our work the average value of the T-O-T angle for each kind of silicon in an aluminosilicate of unknown structure (including amorphous materials) can now be reliably estimated from ^{29}Si -MASNMR alone. This will also facilitate the quantitative interpretation of the spectra of zeolites with more than one crystallographic tetrahedral site.

We thank Professor J. M. Thomas for discussions and encouragement and the BP Research Centre, Sunbury, for support (J.K.) and Exxon Research and Engineering, New Jersey, for an unrestricted grant (S.R.).

Received 28 September 1983; accepted 17 January 1984.

- Lippmaa, E., Magi, M., Samoson, A., Engelhardt, G. & Grimmer, A. R. *J. Am. chem. Soc.* **102**, 4889-4893 (1980).
- Lippmaa, E., Mägi, M., Samoson, A., Tarmak, M. & Engelhardt, G. *J. Am. chem. Soc.* **103**, 4992-4996 (1981).
- Ramdas, S., Thomas, J. M., Klinowski, J., Fyfe, C. A. & Hartman, J. S. *Nature* **292**, 228-230 (1981).
- Fyfe, C. A., Thomas, J. M., Klinowski, J. & Gobbi, G. C. *Angew. Chem. Int. Ed. Engl.* **22**, 259-336 (1983).
- Fyfe, C. A., Gobbi, G. C., Klinowski, J., Thomas, J. M. & Ramdas, S. *Nature* **296**, 530-533 (1982).
- Thomas, J. M., Klinowski, J., Ramdas, S., Hunter, B. K. & Tennakoon, D. T. B. *Chem. Phys. Lett.* **102**, 158-162 (1983).
- Thomas, J. M., Fyfe, C. A., Ramdas, S., Klinowski, J. & Gobbi, G. C. *J. phys. Chem.* **86**, 3061-3064 (1982).
- Jarman, R. H. *JCS chem. Commun.*, 512-513 (1983).
- Grimmer, A. R., von Lampe, F., Tarmak, M. & Lippmaa, E. *Chem. phys. Lett.* **97**, 185-187 (1983).
- Smith, J. V. & Blackwell, C. S. *Nature* **303**, 223-225 (1983).
- Gibbs, G. V., Meagher, E. P., Newton, M. D. & Swanson, D. K. in *Structure and Bonding in Crystals* (eds O'Keeffe, M. & Navrotsky, A.) (Academic, New York, 1981).

- Thomas, J. M., Bursill, L. A., Lodge, E. A., Cheetham A. K. & Fyfe, C. A. *JCS chem. Commun.*, 276-277 (1981).
- Pluth, J. J. & Smith, J. V. *J. Am. chem. Soc.* **102**, 4704-4708 (1980).
- Melchior, M. T., Vaughan, D. E. W., Jarman, R. H. & Jacobson, A. J. *Nature* **298**, 455-456 (1982).
- Klinowski, J., Ramdas, S., Thomas, J. M., Fyfe, C. A. & Hartman, J. S. *JCS Faraday Trans. II* **78**, 1025-1050 (1982).
- Olson, D. H. *J. Phys. Chem.* **74**, 2758-2764 (1970).
- Gallezot, P., Beaumont, R. & Barthomeuf, D. *J. phys. Chem.* **78**, 1550-1553 (1974).
- Klinowski, J., Thomas, J. M., Fyfe, C. A. & Hartman, J. S. *J. phys. Chem.* **85**, 2590-2594 (1981).
- Bresciani Pahor, N., Calligaris, M., Nardin, G. & Randaccio, L. *Acta crystallogr.* **B38**, 893-895 (1982).
- Klinowski, J., Thomas, J. M., Ramdas, S., Fyfe, C. A. & Gobbi, G. C. *2nd Workshop on the Adsorption of Hydrocarbons on Microporous Sorbents*, Eberswalde, DDR, Vol. 2 (Suppl.) (Akademie der Wissenschaften der DDR, 1982).
- Fischer, K. *N. Jahrb. Mineral. Monatsh.* 1-13 (1966).
- Smith, J. V., Knowles, C. R. & Rinaldi, F. *Acta crystallogr.* **17**, 374-384 (1964).
- Engelhardt, G., Mägi, M. & Lippmaa, E. *2nd Workshop on the Adsorption of Hydrocarbons on Microporous Sorbents*, Eberswalde, DDR, Vol. 2, 1 (Akademie der Wissenschaften der DDR, 1982).
- Fälth, L. & Hansen, S. *Acta crystallogr.* **B35**, 1877-1880 (1979).
- Alberti, A. & Vezzadini, G. *Acta crystallogr.* **B37**, 781-788 (1981).
- Schramm, V. & Fischer, K. *Adv. Chem. Ser.* **101**, 259-265 (1971).
- Ferraris, G., Jones, D. W. & Yerkess, J. Z. *Kristallogr.* **135**, 240-252 (1972).
- Olson, D. H., Kokotailo, G. T., Lawton, S. L. & Meier, W. M. *J. phys. Chem.* **85**, 2238-2243 (1981).
- Baerlocher, C. H. & Meier, W. M. *Helv. chim. Acta* **52**, 1853-1860 (1969).
- Galli, E. *Cryst. Str. Commun.* **3**, 339-344 (1974).
- Gard, J. A. & Tait, J. M. *Acta crystallogr.* **B28**, 825-834 (1972).
- Schlenker, J. L., Pluth, J. J. & Smith, J. V. *Mat. Res. Bull.* **14**, 849-856 (1979).
- Klinowski, J., Thomas, J. M., Audier, M., Vasudevan, S., Fyfe, C. A. & Hartman, J. S. *JCS chem. Commun.* 570-571 (1981).
- Parise, J. B. & Cox, D. E. *Z. Kristallogr.* (in the press).

Determination of the Si-O-Si bond angle distribution in vitreous silica by magic angle spinning NMR

E. Dupree & R. F. Pettifer*

Department of Physics, University of Warwick, Coventry CV4 7AL, UK

A necessary condition for any proposed description of the structure of glass is that a model of the glass will reproduce the details of the pair correlation function which may be determined from elastic scattering experiments (X-ray, neutron or electron diffraction). From a knowledge of the structure, excitations (electronic or vibrational) may be deduced. The most extensively studied prototype glass is vitreous SiO_2 . It has long been recognized that the SiO_2 tetrahedra are well preserved in the glass and that it is the interconnection of the tetrahedra which provides the clue to the lack of translational symmetry characteristic of the glassy state. The natural approach has been to build macroscopic models of the structure which satisfy the bond conductivity between atoms whilst preserving the regularity of the SiO_2 tetrahedra. When completed the models are scaled and tested against diffraction data and the macroscopic density. Here, magic angle spinning nuclear magnetic resonance (NMR) spectroscopy has been applied to glassy SiO_2 . The observed line-shape allows the published distribution functions for the Si-O-Si bond angle, crucial in determining the structure of vitreous SiO_2 , to be tested. All of the published distribution functions are inconsistent with our data and a new distribution is proposed.

Many models of vitreous silica have been built¹⁻⁶ and may be characterized principally by the position and shape of the distribution of the first Si-Si distance and an analysis of the shortest ring statistics. These two parameters are not independent but the Si-Si distribution corresponds to the third peak in the radial distribution function (r.d.f.) derived from elastic scattering whereas the ring statistics predominantly influence peaks beyond the third in the r.d.f. All of the above models differ in the peak position, width and asymmetry of the Si-Si distribution as well as their ring statistics.

* Present address: European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, 2000 Hamburg 52, FRG.

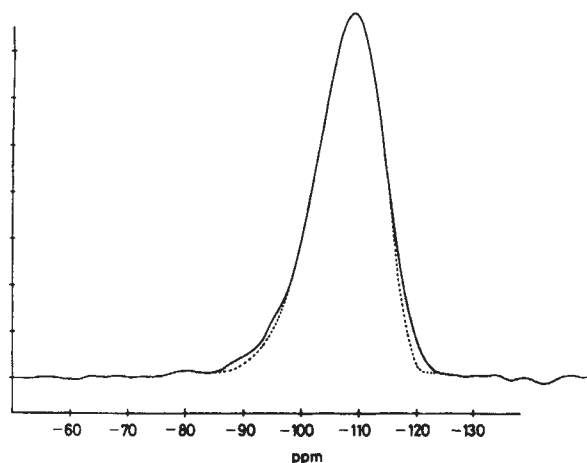


Fig. 1 The measured ^{29}Si line from glassy SiO_2 . The x-axis scale is relative to TMS. The resolution of the spectrometer was 0.2 p.p.m. Also shown is the calculated line (-----) from the assumed distribution function shown with the same annotation in Fig. 2a.

Many elastic scattering experiments have been performed on vitreous SiO_2 (refs 7–11), the most definitive being that of Mozzi and Warren⁷. From their data these authors analysed the third peak of the r.d.f. to produce the Si–Si pair correlation function $G(r)$ and hence from this the bond angle distribution function $V(\alpha)$ (α is the Si–O–Si angle). Mozzi and Warren⁷ succeeded in deducing the third neighbour peak but were less successful for subsequent peaks. The peak in $V(\alpha)$ was found to be at 144° but this did not compare well with the current models and prompted a reappraisal of the original experimental data by Da Silva *et al.*¹⁰ who claimed that $V(\alpha)$ peaked at 153° , more in accord with the model builders. (However, we believe that $V(\alpha)$ calculated in ref. 10 resulted from an incorrect transformation from $G(r)$.) Thus, it is clear that while $V(\alpha)$ is highly model sensitive, doubt exists over the exact form of $V(\alpha)$ from diffraction data. A recent EXAFS measurement of the local structure of silicate glasses claimed to have found a Si–O–Si bond angle of 160° with a static disorder in the bond angle of $\pm 20^\circ$ (ref. 12). It is in this context that we have applied the relatively new technique of magic angle spinning NMR to determine $V(\alpha)$ directly.

This technique removes the static dipolar broadening normally present in solids, and in crystalline materials allows a determination of the local structural order similar to that obtained by high resolution NMR in liquids. It has recently been shown, for instance, that SiO_4 tetrahedra linked in different ways have very different chemical shifts^{13,14}. Recently Smith and Blackwell¹⁵ have shown that in different silica polymorphs there is a linear correlation between the chemical shift δ and the secant of the Si–O–Si angle α , with a correlation 0.9981 and a less good correlation between the shift and the Si–Si distance. Some theoretical support for the correlation between shift and $\sec \alpha$ comes from the calculations of Newton and Gibbs¹⁶ who showed that the bond overlap populations vary linearly with $\sec \alpha$ over the range of at least 130° – 180° in various silicic acid molecules and in coesite, whilst the chemical shift is also expected to vary linearly with bond overlap¹⁷.

The magic angle spinning NMR data give the distribution function $W(\beta)$ of the mean secant of the Si–O–Si bond angles. To make contact with diffraction studies, the bond angle distribution function $V(\alpha)$ can be determined from the pair correlation function $G(r)$ using the relation

$$V(\alpha) = G(r) \left| \frac{dr}{d\alpha} \right| \quad (1)$$

and $V(\alpha)$ can in turn be transformed into a distribution function

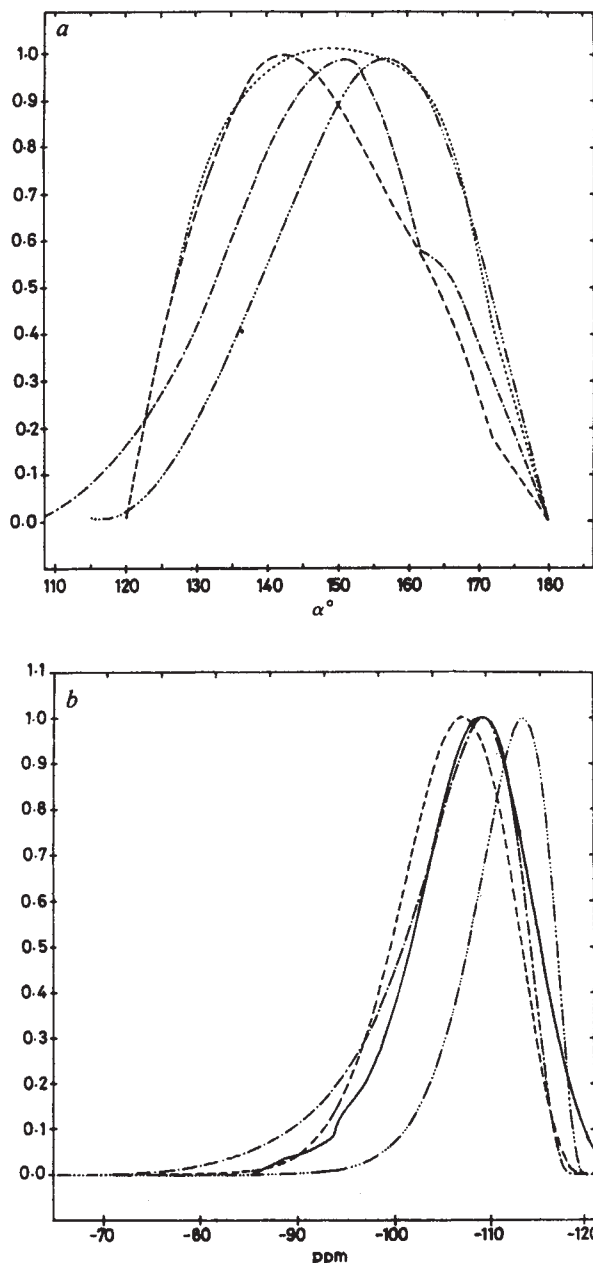


Fig. 2 a, Four bond angle distribution functions which most closely match the NMR data; —, Evans and Teter⁴; ----, Mitra⁶; —., Mozzi and Warren⁷; -.-.-, an assumed distribution function which matches the NMR data shown in Fig. 1. As some of these distributions possess large tails which are difficult to estimate, all curves have been scaled to have unit height at their maxima. b, The corresponding calculated NMR line shapes for the first three distributions in a together with the experimental line shape (—).

in terms of the random variable β , where $\beta = \sec \alpha$

$$X(\beta) = V(\alpha) \left| \frac{d\alpha}{d\beta} \right| \quad (2)$$

Note here that $X(\beta)$ is the distribution function of all secants of the Si–O–Si bond angle whereas $W(\beta)$ is the distribution of the mean secant per silicon tetrahedron. To compare $X(\beta)$ with $W(\beta)$ we assume that the individual Si–O–Si bond angles at each corner of the tetrahedral unit are statistically independent, an assumption perfectly consistent with that implicit in the model building approach. The justification is that the bond angle distribution is dominated by the topology of the network. Thus, the characteristic function of $W(\beta)$ (that is, the Fourier transform

(FT) of $W(\beta)$ is given by

$$\text{FT}\{W(\beta)\} = \text{FT}\left\{W\left(\frac{\beta_1 + \beta_2 + \beta_3 + \beta_4}{4}\right)\right\} \quad (3)$$

with

$$\text{FT}\{W(\beta_1 + \beta_2 + \beta_3 + \beta_4)\} = (\text{FT}\{X(\beta)\})^4 \quad (4)$$

where FT{} denotes Fourier transform and β_{1-n} are the secants of the bond angles α_{1-4} at the four corners of the tetrahedron. Using the convolution theorem of Fourier transforms, this is equivalent to convolving four $X(\beta)$ distributions and rescaling the independent variable by a factor of four. It is well known that arbitrary distributions when convolved rapidly evolve into gaussian functions, so such self convolution will produce a loss of sensitivity to the original distribution $V(\alpha)$; nevertheless, the mean and variance of the distribution are preserved and after three convolutions asymmetry in the NMR line-shape is still visible.

The NMR spectrum was obtained on a Bruker CXP-300 spectrometer operating at 59.6 MHz for the ^{29}Si resonance. The spinning rate was ~ 3.5 kHz. Owing to the relatively long relaxation time of the ^{29}Si resonance, a 10-s delay between pulses was used together with a $\pi/8$ pulse. The sample was Heraeus silica coarsely crushed into a powder, and the spectrum obtained after 2,118 pulses is shown in Fig. 1. Note that the line-shape is anisotropic, with a steeper fall on the large shift side arising from the bond angle distribution. Figure 2a shows the bond angle distribution deduced by Mozzi and Warren⁷ together with the distribution obtained in a recent molecular dynamics calculation for SiO_2 (ref. 6) and by the theoretical model which gives closest agreement with the experimental NMR line-shape⁴. Figure 2b shows the NMR line-shape calculated from each of these $V(\alpha)$ values. It is clear that the calculated line-shape and position are sensitive to the details of the $V(\alpha)$ used. The theoretical distribution which predicts an NMR line closest to experiment is the 'relaxed' model of Evans and Teter⁴. However, even though this predicts approximately the correct position for the centre of the line, the shape deviates markedly from that observed. In particular the small shift side of the line is enhanced and the large shift side drops more rapidly than experiment. We believe this is because the Evans and Teter model has too large a $V(\alpha)$ for small bond angles ($\leq 120^\circ$) and too small a $V(\alpha)$ for large bond angles ($> 160^\circ$). A $V(\alpha)$ which gives excellent agreement with experiment is shown in Fig. 2a and the fit obtained in Fig. 1. This has a very broad distribution of angles with roughly constant probability over 140 – 155° and a full width at half maximum larger than predicted by the various models.

The SCF-MO calculations of Newton and Gibbs¹⁶ find an equilibrium bond angle of $\sim 140^\circ$ with a shallow minimum which rises more steeply in the low α region. Our $V(\alpha)$ is slightly skewed, with the steep side of the distribution on the low α side consistent with this calculation.

Despite the continued interest in model building it is clear that all models are still some way from good agreement with experiment. This study should provide the impetus for a quantitative comparison of models because our data, although insufficient to determine the structure uniquely, give an exacting constraint on these models.

We thank Bruker Spectrospin and Dr H. Forster for his help in obtaining the spectra, and the University of Warwick for support.

Received 4 November 1983; accepted 3 January 1984.

1. Ordway, F. *Science* **143**, 800–801 (1964).
2. Bell, R. J. & Dean, P. *Phil. Mag.* **25**, 1381–1398 (1972).
3. Evans, D. L. & King, S. V. *Nature* **212**, 1353–1354 (1966).
4. Evans, D. L. & Teter, M. in *The Structure of Non-Crystalline Solids* (ed. Gaskell, P. H.) 53–57 (Taylor & Francis, London, 1977).
5. Ching, W. Y. *Phys. Rev.* **26**, 6610–6621 (1982).
6. Mitra, S. K. *Phil. Mag.* **45B**, 529–548 (1982).
7. Mozzi, R. L. & Warren, B. E. *J. appl. Crystallogr.* **2**, 164–172 (1969).
8. Henninger, E. H., Bushart, R. C. & Heaton, L. *J. phys. Chem. Solids* **28**, 423–432 (1967).
9. Konner, J. H. & Karle, J. *Acta crystallogr.* **29A**, 702–710 (1973).

10. Da Silva, J. R. G., Pinatti, D. G., Anderson, C. E. & Rudee, M. L. *Phil. Mag.* **31**, 713–717 (1975).
11. Wright, A. C. & Leadbetter, A. J. *Phys. Chem. Glasses* **17**, 122–145 (1976).
12. Greaves, G. N., Fontaine, A., Lagarde, P., Roux, D. & Gurman, S. J. *Nature* **293**, 611–616 (1981).
13. Lippmaa, E., Magi, M., Samoson, A., Engelhardt, G. & Grimmer, A. R. *J. Am. chem. Soc.* **102**, 4889–4893 (1980).
14. Klinowski, J., Thomas, J. M., Fyfe, C. A. & Hartmann, J. S. *J. phys. Chem.* **85**, 2590–2594 (1980).
15. Smith, J. V. & Scott Blackwell, C. *Nature* **303**, 223–225 (1983).
16. Newton, M. D. & Gibbs, G. V. *Phys. Chem. Minerals* **6**, 221–246 (1980).
17. Harris, R. K. & Mann, B. E. (eds) *N.M.R. and the Periodic Table* (Academic, New York, 1978).

Natural calcite in cathodoluminescence: crystal growth during diagenesis

G. M. Walkden & J. R. Berry

Department of Geology, University of Aberdeen, Marischal College, Aberdeen AB9 1AS, UK

Some of the largest crystals found as pore-filling cements in ancient limestones are those which develop as single-crystal (syntaxial) overgrowths on crinoid grains. Although originally porous and polycrystalline^{1,2}, such grains normally behave as monocrystalline substrates during diagenesis, facilitating the relatively rapid growth of optically contiguous calcite haloes. The crystal form and growth morphology of pore-fill cements can yield important information about the diagenetic environment¹, but growth features cannot easily be detected with the light microscope. Chemical stains have been used to study internal growth zonation^{3,4} but cathodoluminescence⁵ has proved a more satisfactory technique^{6,7} and our own studies of calcite have revealed a variety of novel growth phenomena. We distinguish four styles of syntaxial overgrowth, the morphologies of which appear to depend on three main factors: grain surface conditions, surrounding pore conditions and the nature of the diagenetic environment.

In a fresh state the solid framework (stereom system) of crinoid and other echinoderm grains consists of high-Mg calcite^{1,2}. This mineral phase rarely survives unchanged during diagenesis^{2,8} and syntaxial overgrowths apparently always consist of low-Mg calcite^{9–11}. Evamy and Shearman^{3,4} established that crinoid syntaxial overgrowths develop fastest in *c*-axial directions starting with the seeding of numerous small discrete crystal spires on the host grain. Through enlargement with growth these progressively coalesce to produce fewer and larger terminations (Fig. 1a). Overgrowths of this type are persistently euhedral and we have termed them 'spired overgrowths' (Figs 1a, 2a), but in many rocks this style is rare or absent. Figure 2a shows an example from a Carboniferous limestone turbidite originally deposited in a marine basinal setting. The spires may well be seeded on the outwardly directed pore walls of the host grain and they evidently create a crystallographically uniform pellicle round the grain by bridging pores and other inhomogeneities.

'Needle overgrowths' (Figs 1b, 2b) are a rare special case of spired overgrowths where the initial fringe has failed to undergo amalgamation with growth, and instead the individual spires have perpetuated themselves as discrete parallel needles. The intergranular areas of this rock contain much non-carbonate mud entrapped between cement crystals and separating the needles of the overgrowth (Fig. 2b). It was evidently the entrainment of this mud that inhibited coalescence leading to the perpetuation of the acicular form.

Despite the persuasive theoretical basis³ for assuming that overgrowth development starts with a spired stage, there are many crinoidal limestones in which early spired overgrowths are not characteristic. 'Contouring overgrowths' (Figs 1c, 2c) begin with patchily distributed contouring or irregular growths in *c*-axial directions: spires are extremely rare and never grow

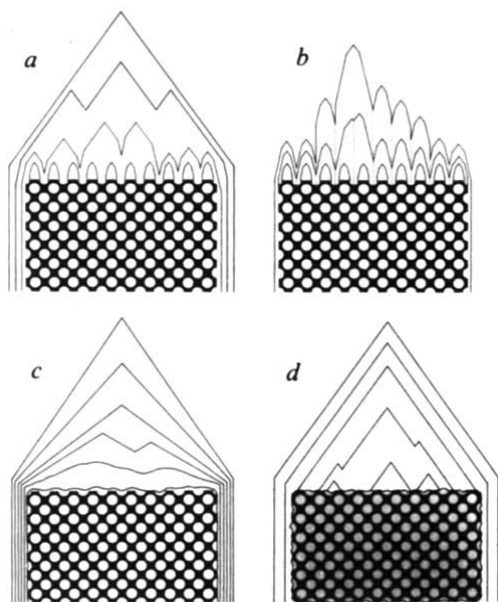


Fig. 1 Schematic representation of four distinct syntaxial overgrowth styles on echinoderm grain substrates. *a*, Spired overgrowth; *b*, needle overgrowth; *c*, contouring overgrowth; *d*, idiomorphic overgrowth. All patterns represent growth in the *c*-axial direction.

beyond an incipient stage. Subsequent development perpetuates this subdued morphology, and growth surfaces are commonly curved or fretted (Fig. 2*c*). Examples have been illustrated by Freeman⁶, Meyers^{7,12} and Walkden and Berry⁹, and in each case a shallow meteoric diagenetic environment has been deduced on the basis of several criteria. Our example (Fig. 2*c*) comes from a former cyclic shallow marine carbonate platform setting in northern England. Euhedral patterns seldom appear until late and growth zonation commonly reveals former growth surfaces which, because of irregularity or geometric relationships, could not correspond to true crystal faces. These unusual growth forms are discussed separately.

'Idiomorphic overgrowths' (Figs 1*d*, 2*d*) display zones that reflect the very early establishment of perfect or near-perfect crystal faces. Spires are again never developed, and growth of faces is by regular parallel increments. This style is uncommon in our experience and may reflect unusual circumstances: our example (Fig. 1*d*) comes from a late Triassic deposit made largely of crinoid debris derived from the weathering of nearby outcrops of older limestone. These grains had already undergone a cycle of burial and cementation before deposition, and their pore systems were still filled with an original Carboniferous cement when renewed cementation took place.

Each of the four examples we show reflects a distinct set of diagenetic circumstances that may have profoundly influenced crystal development. Spired overgrowths (Figs 1*a*, 2*a*) are not the predominant style, but their persistently euhedral morphology may reflect nearly ideal growth conditions following seeding upon a structurally discontinuous fresh grain surface. We have suggested elsewhere⁹ that the rocks containing our example were initially cemented in a comparatively stable diagenetic environment not far removed from the original marine basinal depositional setting. By contrast, the evidently stunted morphology of the contouring overgrowths (Figs 1*c*, 2*c*) implies inhibiting factors, and fretted growth zones, interpreted as dissolution effects within a meteoric diagenetic environment^{9,12}, demonstrate one possible spire-inhibiting factor. Another may have been surface tension effects within the vadose zone⁶. Surface homogenization was probably achieved instead through cement growth within the pore systems. Like the spired overgrowths, idiomorphic overgrowths (Figs 1*d*, 2*d*) also reflect nearly ideal growth conditions, but in this style early spire growth

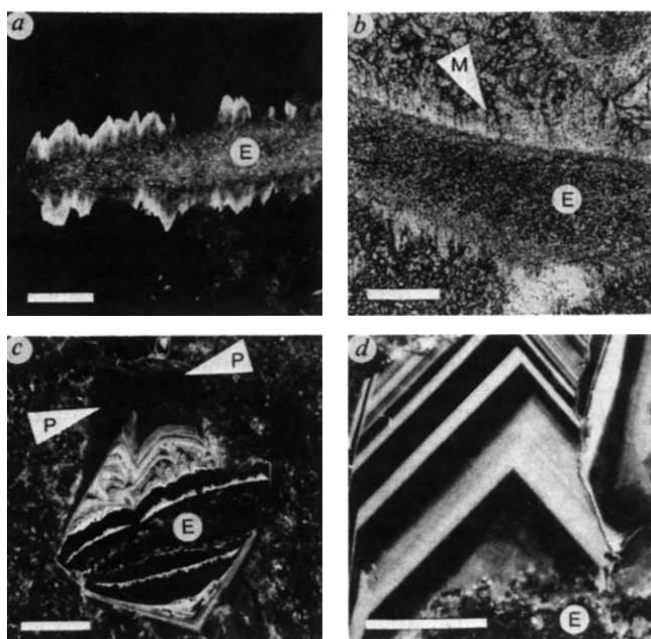


Fig. 2 Photomicrographs of the syntaxial overgrowth styles shown in Fig. 1. E, Echinoderm grains. Scale bar, 0.2 mm. *a*, Spired overgrowth in Lower Carboniferous limestone turbidite, Craven Basin, northern England. Spire amalgamation at a semi-mature stage (cathodoluminescence). *b*, Needle overgrowth in Devonian limestone, New York State, USA. Needle-like spire growths are separated by trails of mud inclusions (M) (plane light). *c*, Contouring overgrowth in Lower Carboniferous limestone, Derbyshire, England. Bright cements overlie erosional hiatuses. Late stages show unusual non-facial planar growth surfaces (P) (cathodoluminescence). *d*, Idiomorphic overgrowth on Carboniferous echinoderm grain in Triassic cemented limestone, Cotswolds, England. Well-organized parallel growth of faces (cathodoluminescence).

could have been unnecessary because the original pore system was already homogenized by an earlier cement. Needle overgrowths (Figs 1*b*, 2*b*) are a special case of spired overgrowth where amalgamation was prevented by entrainment of impurities. They reflect growth by displacement or replacement into impure interstitial carbonate mud.

Cathodoluminescence of diagenetic calcites has revealed a number of unusual growth phenomena which may provide new information on the ways in which crystals grow in diagenetic conditions. 'Sector zoning'¹³ may account for regularly shaped patches of distinct luminescence affecting one specific face or a particular sector of a crystal (Fig. 3*b*). More rarely we have noted concentrations of abnormal luminescence colour such as blue or green within normally orange calcite. We interpret these effects as the result of differential uptake of luminescence activator ions at specific sites determined by the sector-zoning phenomenon¹³, and the effect is common where growth has been rapid such as in areas of repair. Too much uptake of impurity ions may disrupt crystal growth altogether and may account for areas of confused and chaotic growth (Fig. 2*c*).

'Pseudofaces' are a new phenomenon recognized where successive planar growth zones, normally interpretable as prograding face growth, have undergone a progressive change in attitude (Figs 1*c*, 2*c*, 3*a*). They are particularly common in the late stages of contouring overgrowths. Such hinging growth planes could never be assigned an integrated set of workable Miller indices and their relationships seem to conflict with a basic law of crystallography whereby faces should grow through the addition of parallel increments. These must therefore represent some sort of non-facial growth surface and they are termed pseudofaces.

Despite this lack of crystallographic credibility, however, groups of pseudofaces appear to have evolved synchronously.

Fig. 3 Overgrowths of the contouring type on echinoderm grains (E) from the Lower Carboniferous limestone of Derbyshire, England (cathodoluminescence). Scale bar, 0.2 mm. *a*, Overgrowth showing solution hiatuses (S) and pseudofaces (P). Pseudofaces show a progressive change in attitude. The outermost etched surface displays localized rapid repair (R). *b*, Early and middle stages of overgrowth show irregular (I) and pseudoface (P) development and later stages display form changes (C) with normal parallel face development. Sector zoning is visible at Z.

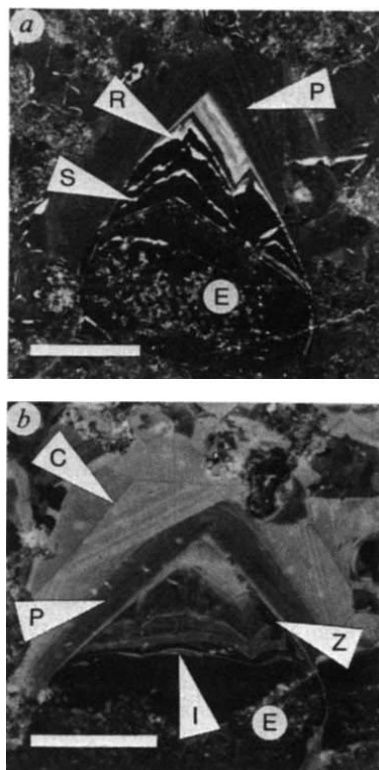


Figure 2c represents a section through a calcite crystal cut obliquely through its *c*-axis so that all three rhombohedral planes comprising one termination are intersected. Growth zones are confluent across these but only two planes show any progressive angular change; growth zones on the third are perfectly parallel. This relationship demonstrates that angular change only involved intercepts on the *c*-axis, and that the rhombohedron was becoming more acute. The opposite termination evolved in an identical manner (Fig. 2c), and it can be concluded that, like true crystal faces, pseudofaces can comprise a symmetry-related group evolving in a coordinated manner.

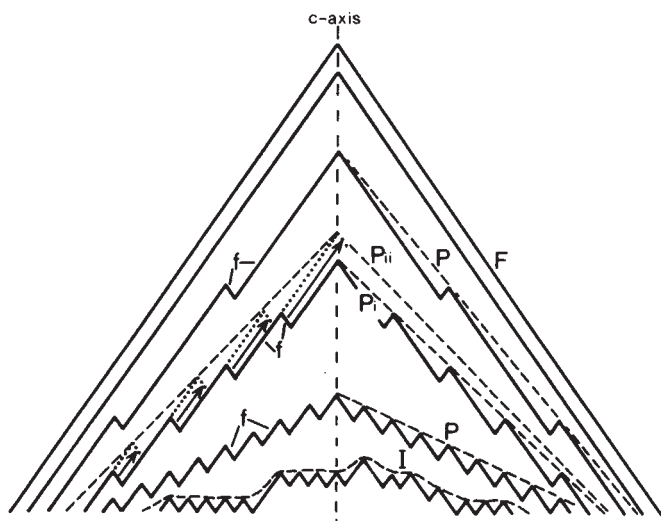


Fig. 4 Evolution of crystal faces from an originally irregular crystal substrate (I) by the intermediate development of planar pseudofaces. This schematic diagram shows how small independently growing facets (f), on a more reduced scale than shown, could define the positions of a series of planar non-facial growth surfaces or pseudofaces (P). In this simplified situation the growth of facets on any one pseudoface is unidirectional and the rate of growth progressively increases towards the apex, as represented by the length of the arrows. Such growth would produce a successive change in the attitude of each pseudoface until a crystal face position is encountered where normal parallel progradation may take over (F).

Pseudofaces thus appear to be an unconventional means of achieving a change in crystal habit through the production of many intermediate stages rather than by the normal method involving the overdevelopment or underdevelopment of the faces of one crystal form relative to another (Fig. 3b). In Fig. 3a pseudoface evolution seems to have started early with the appearance of an obtuse rhombohedral form and by the end of growth this has evolved progressively into an acute rhombohedron. By contrast, in Fig. 3b there has been oscillatory development between two competing forms but the changes have been achieved mostly through normal parallel development.

To account for the apparently non-crystallographic but nevertheless organized development of pseudofaces, we suggest that they might be constructed of many submicroscopic facets which do correspond to actual crystal faces (f in Fig. 4). These facets would necessarily be arranged in a regular stepped pattern, the precise step spacing determining the orientation of the pseudoface (Fig. 4, P). Progressive angular change of such a pseudoface could then be achieved by a regular increase in growth rate of facets towards the position of the *c*-axis (Fig. 4, Pi, Pii). The implicit crystallographic control over positioning and rate of growth of the facets might also be responsible for ensuring that groups of pseudofaces developed synchronously about a centre of symmetry. Should inhibiting factors prevent crystallographically controlled pseudoface growth then curved or irregular surfaces may develop (Fig. 4, I).

We thank Peter Brown, Ian Parsons and Lesley Glasser for their useful advice, and Robin Bathurst and Bill Meyers gave welcome comments and guidance. We acknowledge the financial support of the NERC and the Royal Society.

Received 18 July 1983; accepted 17 January 1984.

1. Towe, K. M. *Science* **157**, 1048–1050 (1967).
2. Bathurst, R. G. C. *Carbonate Sediments and their Diagenesis* (Elsevier, Amsterdam, 1975).
3. Evamy, B. D. & Shearman, D. J. *Sedimentology* **5**, 211–233 (1965).
4. Evamy, B. D. & Shearman, D. J. *Sedimentology* **12**, 317–322 (1969).
5. Sippel, R. F. & Glover, E. D. *Science* **150**, 1283–1287 (1965).
6. Freeman, T. *Nat. phys. Sci.* **233**, 133–134 (1971).
7. Meyers, W. J. *J. sedim. Petrol.* **44**, 837–861 (1974).
8. Longman, M. W. *Bull. Am. Ass. Petrol. Geol.* **64**, 461–487 (1980).
9. Walkden, G. M. & Berry, J. R. *Sedimentology* **31** (in the press).
10. Lohmann, K. C. & Meyers, W. J. *J. sedim. Petrol.* **47**, 1078–1088 (1977).
11. Meyers, W. J. & Lohmann, K. C. *J. sedim. Petrol.* **48**, 475–488 (1978).
12. Meyers, W. J. *Sedimentology* **25**, 371–400 (1978).
13. Dowty, E. *Am. Miner.* **61**, 460–469 (1976).

Fanning of fracture zones and a three-dimensional model of the Mid-Atlantic Ridge

W. R. Roest*, R. C. Searle† & B. J. Collette*

* Vening Meinesz Laboratorium, afdeling Mariene Geofysica, Budapestlaan 4, PO Box 80.021, 3508 TA Utrecht, The Netherlands
† Institute of Oceanographic Sciences, Brook Road, Wormley, Godalming, Surrey GU8 5UB, UK

Present-day spreading directions as represented by the active (transform fault) sections of fracture zones in the central North Atlantic do not fit a small circle pattern about a unique pole^{1–5}, as recently confirmed by a 'Gloria' survey. Contrary to this pattern, which is geometrically required if ocean floor spreading is to occur⁶, the active transform fault directions show a convergence towards the east³. Because of the fanning of the present directions, the African and American plates are caught in what might be likened to a dovetail construction. To allow motion, the African plate must shrink about 2% in a north-south direction. Such a shrinkage might be accomplished by thermal contraction. Tension due to the horizontal component of this contraction has been proposed as the primary cause of fracture zones^{4,7}. However, horizontal thermal contraction alone cannot explain the recorded pattern of spreading segments and transform faults, because it does not explain why transform faults

converge to the east and diverge to the west. For this, an active driving mechanism is needed. We show here that the gravitational drive (ridge push) caused by thermal contraction in the lithosphere formed along a sinusoidal mid-ocean ridge delivers such a mechanism.

To check the fanning of transform fault directions, the Institute of Oceanographic Sciences organized a Gloria⁸ long-range sidescan sonar survey in autumn 1981 onboard MV *Farnella*. The survey covered the active sections of all major fracture zones (Oceanographer, Hayes, Atlantis, Kane and Fifteen Twenty) on the Mid-Atlantic Ridge between 35° N and 15° N. Significant lineaments within the fracture zones were traced from 1:250,000 scale sonar images, and corrected for slant-range distortion.

The Gloria results show in each fracture zone a band, 2–5 km wide, of 2–36-km-long lineaments striking approximately parallel to the direction of the fracture zone (Fig. 1). These bands are believed to represent the zones of active transform faulting. The strikes of the individual lineaments vary by up to 5° on either side of the mean trend. Table 1 lists the overall trend of the transform zones, the range of trends, their means and the standard error of these means, the predicted transform fault directions from the RM2 pole by Minster and Jordan⁹ and the differences from these latter directions. The zones of active transform faulting as defined above converge to the east and do not fit the RM2 pole nor any other realistic pole¹⁰. The deviation of the overall directions of the fracture zones with respect to the RM2 pole is systematic and varies from about +3° for Oceanographer and Hayes Fracture Zones to –4° for the Fifteen Twenty Fracture Zone (see also Fig. 2).

The Fifteen Twenty Zone was surveyed in more detail in 1982 by the Vening Meinesz Laboratorium onboard MV *Tyro*. This survey showed that the transform zone is in a leaky mode. To the north and the south the present spreading direction is 090.5°. The fracture zone valley forms a rhombochasm with sides varying from 090.5° to 099°. The 095.5° direction defined by the Gloria data might be interpreted as the active transform faulting direction, or the grain of a possibly intrusive body which occurs along the longer diagonal of the rhombochasm. In the latter case the active transform fault directions would vary from

090.5° near the spreading axis to 099° at a distance of 195 km from this axis. The first direction is parallel to that of the Marathon Fracture Zone at 12°40' N, which has an offset of 80 km and whose direction is well defined by a narrow valley only a few kilometres wide¹¹. The varying transform fault directions in the Fifteen Twenty Fracture Zone are also reflected in the seafloor spreading topography to the north and south.

The occurrence of fracture zone directions which differ from the present transform fault directions is generally interpreted as the result of a change in seafloor spreading direction^{4,5,12,13}. No consensus exists, however, with regard to the date of this change, which varies from 9 to 3 Myr ago. Combining this with the circumstance that the present directions do not make sense in a rigorous geometric spreading model, we suggest that the varying directions reflect the process of thermal contraction in the horizontal plane, that is, a gradual adjustment of the local spreading direction to the general direction of plate movement.

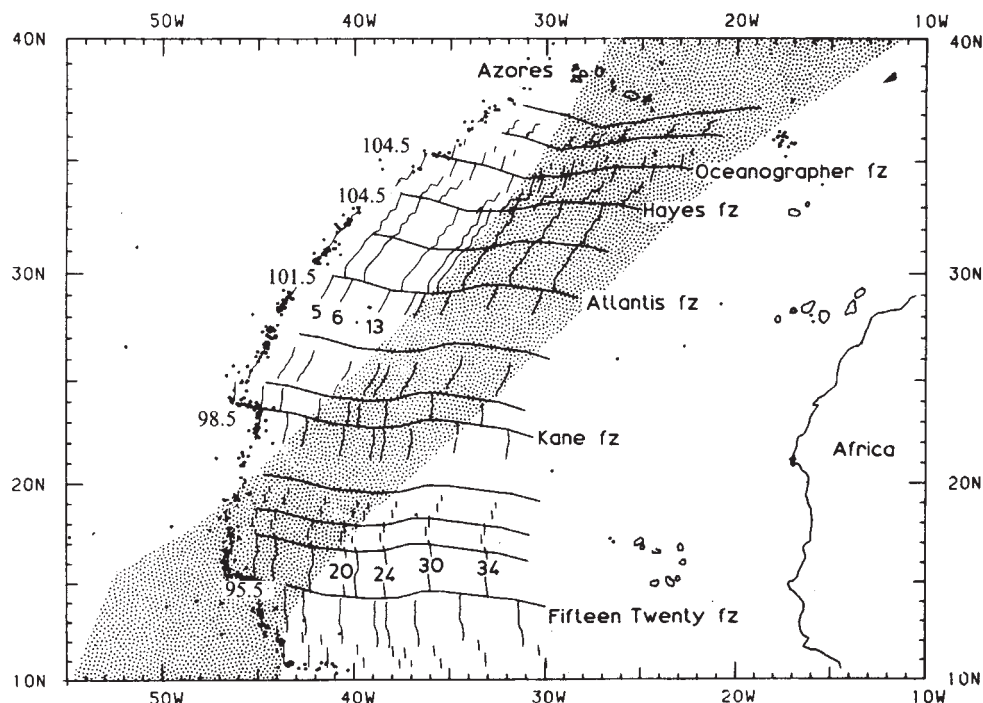
A preliminary analysis¹⁴ of the fossil fracture zone pattern east of the Mid-Atlantic Ridge between magnetic anomalies 5 and 34, as reconnoitred during the Kroonvlag project, led to the construction of the flowline pattern of Fig. 2. These flowlines describe the history of the separation of the American and African plates from 84 to 9.5 Myr ago. For the analysis, both seismic and magnetic data were used along 39 tracks between the English Channel and the north coast of South America¹⁰. No fanning of fracture zones can be observed for lithosphere older than anomaly 13 (~35 Myr¹⁵). For oceanic lithosphere younger than anomaly 13, however, the situation is different: there, fracture zone directions converge towards the east, although less than observed in the transform domains.

The fact that the fanning of fracture zone directions is restricted to the younger part of the ocean floor indicates a relationship with the lithospheric cooling process. Both the relationship between lithospheric age and oceanic depth^{16–18}, and the ridge push as one of the driving forces in plate tectonics¹⁹, are related to the cooling process of new oceanic lithosphere. By taking the horizontal component of thermal contraction into account, we will show that the amount of shortening of the African plate needed to free the plates from the dovetail construction and



Fig. 1 Tectonic lineaments inferred from Gloria sonographs of Kane Fracture Zone. Heavy lines, principle lineaments of the transform zone; medium lines, other lineaments; light broken lines, ship's track; light continuous lines, outline of sonograph coverage; dotted lines, isobaths (in km) from ref. 26; stipple, areas <3,000 m deep.

Fig. 2 Azimuths of major fracture zones in the central North Atlantic based on a Gloria survey conducted by MV *Farnella* in 1981, shown together with preliminary synthetic flow lines and isochrons mainly based on results of the Kroonvlag project (shaded area). Dots are earthquake epicentres.



hence the degree of fanning of the transform faults can also be related to this cooling process.

In the model we propose to account for fanning, the key factor is the ridge push²⁰. Due to the curvilinear shape of the Mid-Atlantic Ridge, the stress trajectories do not form a system of meridians and parallels but fan with the shape of the ridge. On the concave side of the ridge the spreading ocean floor converges to a degree that is determined by the horizontal contraction of the cooling lithosphere as it moves away from the ridge. A complementary divergence occurs on the convex side.

We first investigate whether horizontal thermal contraction can be large enough to account for the observed angular deviations. For this, the depth- and age-dependent temperature was calculated for a two-dimensional thermal plate model¹⁷. Instead of the isotherms usually shown, Fig. 3 gives the temperature as a function of the lithospheric age (t) for different depths z in the 100 km-thick plate. We now introduce the y -direction perpendicular to the direction of spreading and allow free contraction in this direction. This contraction is then immediately related to the temperature distribution in the t - z plane. Using a thermal volume expansion coefficient of $4.5 \times 10^{-5} \text{ } ^\circ\text{C}^{-1}$ and a temperature at the bottom of the plate $T_m = 1,200 \text{ } ^\circ\text{C}$, the relative contraction at the surface is 1.8% (Fig. 3). The angular deviation θ for constant z is defined as the arctangent of the derivative with respect to x ($=vt$) of the absolute horizontal contraction of a ridge segment with a length D . Clearly, θ is a function of the depth in the plate, the lithospheric age, but also of the spreading velocity v and the actual length of the spreading segment D . The half spreading velocity was set at 1.5 cm s^{-1} and the length D at 150 km. As the plate model is not valid

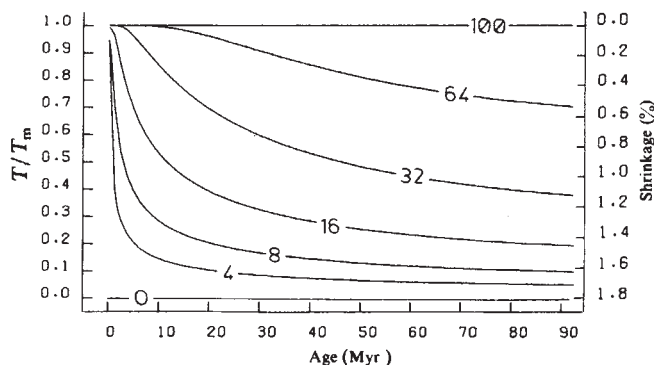


Fig. 3 Temperature and horizontal thermal contraction as a function of lithospheric age, calculated for the thermal plate model¹⁷, at different depths (in km) in the plate.

close to the ridge, a maximum value of θ was estimated by averaging the effect over a mean offset value of 100 km. This θ is then 0.9° at a depth of 8 km and 1.2° at a depth of 4 km. The cumulative effect θ_{cum} of a staircase model of 10 such segments with offsets of 100 km would then vary between 9 and 12° , a value of the right order of magnitude (see Table 1). For older crust, the cumulative effect, θ_{cum} , falls. For ages of 10 and 20 yr, maxima occur of 2.3° at a depth of 16 km, and 1.2° at a depth of 30 km. For crust older than 35 yr θ_{cum} becomes negligible.

To calculate the effect of curvilinear shape of the Mid-Atlantic Ridge on the direction of the ridge push, the temperature distribution was calculated for a three-dimensional sinusoidal model. The solution of the heat equation was approximated by

Table 1 Observed and predicted transform fault directions

Fracture zones	Overall trend of transform zone	Range of trends	Mean trend	Standard error of mean	Predicted from RM2 pole*	Difference with overall trend
Oceanographer	104.5 (107.0)	99–110	107.1	0.7	101.6	+2.9 (+5.4)
Hayes	104.5	96–108	103.2	0.8	101.3	+3.2
Atlantis	101.5	92–108	100.4	0.8	100.8	+0.7
Kane	98.5	88–107	99.3	0.7	100.1	−1.6
Fifteen Twenty	95.5 (90.5)	91–102	95.7	0.5	99.6	−4.1 (−9.1)

Values in italics represent local spreading directions from other sources^{21–24} and unpublished data.

* RM2 pole of Minster and Jordan⁹, situated at 80.4° N , 56.4° E .

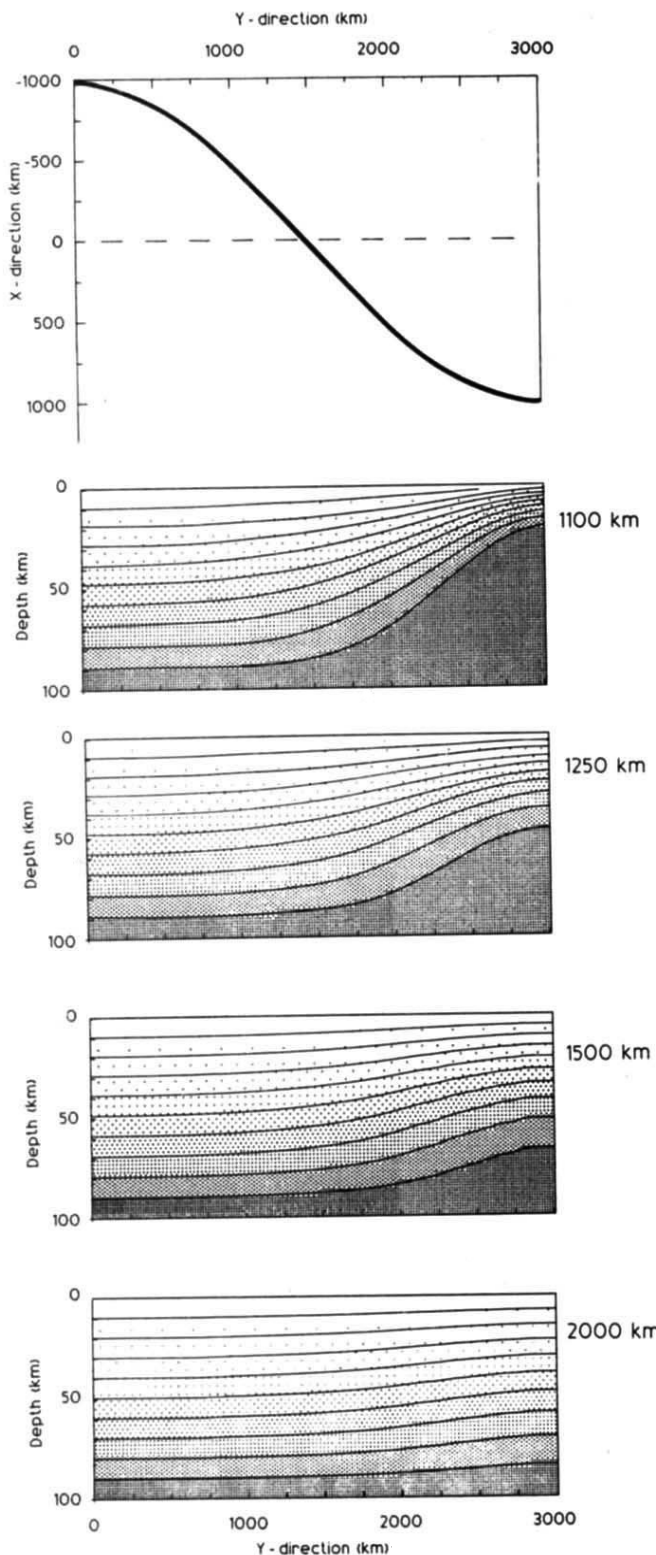


Fig. 4 Distribution of isotherms in vertical cross-sections perpendicular to the direction of spreading, at different distances from $x = 0$. The sinusoidal mid-ocean ridge is shown in the upper part of the figure in a bird's-eye view.

means of a finite difference scheme which has been tested for convergence and stability. A 50×50 km grid was used. Figure 4 shows the result for a sinusoidal ridge of wavelength 6,000 km and amplitude 1,000 km. Isotherms of the steady-state solution in vertical sections perpendicular to the spreading direction are shown at distances of 1,100, 1,250, 1,500 and 2,000 km from the y-axis. The effect of the sinuosity is clearly reflected in the isotherms, becoming less with increasing distance from the ridge. Additionally, the horizontal heat flow is influenced by the shape

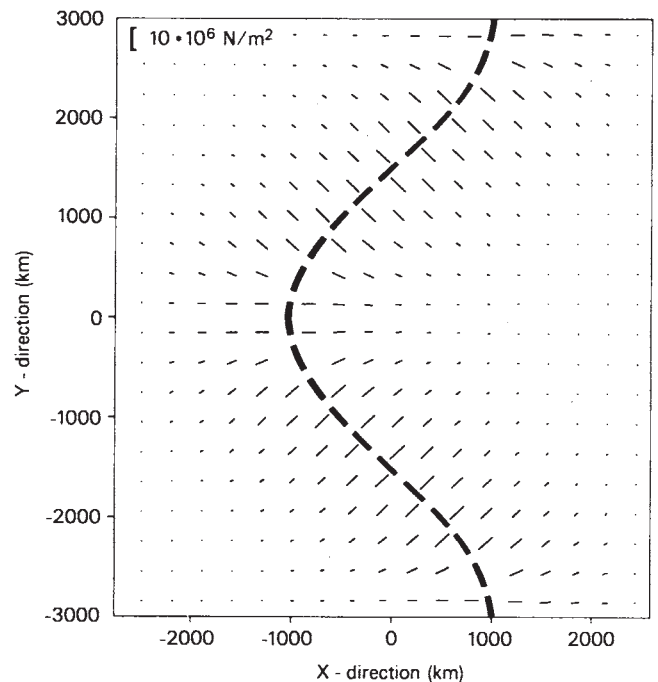


Fig. 5 Direction and amplitude of the incremental ridge-push for the sinusoidal mid-ocean ridge shown in Fig. 4.

of the ridge, but as in the case of two-dimensional models, this horizontal heat flow is very small.

The total ridge push caused by the density distribution in a lithospheric plate is the sum of body forces which are working on that part of the lithosphere that lies between the spreading centre and a subduction zone or continental margin²⁰. It can be calculated by integrating the incremental ridge push along the direction of spreading. The incremental ridge push is the horizontal gradient of the vertically integrated lithostatic pressure.

For the computation of the incremental ridge push, we define the thickness of the lithosphere by the isotherm at its base T_s . Then the two-dimensional horizontal gradient is computed at the grid points of the thermal model. Figure 5 shows the results of the calculations for $T_s = 0.9 \times T_m$, where direction and magnitude of the ridge push are given at discrete points. The maximum angle between ridge push and spreading direction is about 45° , considerably larger than θ_{cum} , and therefore large enough to account for the observed fanning, the actual amount of which is restricted by the space created by the horizontal thermal contraction.

We will clarify this latter remark. The reason that the convergence of the gradient of the lithostatic pressure does not lead to further deformations of the oceanic lithosphere and only takes up the space that becomes free through the cooling process, lies in the fact that the stresses involved are of the order of $150 \times 10^5 \text{ N m}^{-2}$, which is too small to give anelastic deformation, that is, plastic flow or faulting. This means that contraction in the horizontal plane is the limiting factor for the convergence of the fracture zones and, since the divergence of the lithosphere on the convex side of the ridge is complementary to the convergence at the concave side, also for the divergence of fracture zones at the opposite side of the ridge.

The fanning model ends the dispute about the different poles of rotation of the Atlantis and Kane Fracture Zones^{1-3,5}. The circumstance that the Atlantis Fracture Zone shows only one pole of rotation for the past 35–40 Myr¹ results from the fact that it is situated near to the axis of symmetry of the central North Atlantic when considered as a bight. Therefore, it is the only fracture zone whose meridian goes through the pole of spreading and as such can be used to help to determine this pole.

Finally, the fanning of the Atlantic fracture zones leads to the remarkable situation that the distances between fracture

zones on the convex side of the ridge are larger than on the concave side, that is, the isochrons to both sides of the axis will not fit exactly when brought together. This situation probably did not exist in the early stage of opening of the Atlantic. Only later had the relief due to cooling developed sufficiently to produce a substantial ridge push. Continued field studies are needed to verify this and to give us the experimental data to refine the proposed model.

Received 27 June 1983; accepted 9 January 1984.

1. Phillips, J. D. & Luyendyk, B. P. *Science* **170**, 727-729 (1970).
2. Pitman, W. C. III & Talwani, M. *Science* **174**, 845-846 (1971).
3. Phillips, J. D., Luyendyk, B. P. & Forsyth, D. W. *Science* **174**, 846 (1971).
4. Collette, B. J. *Nature* **251**, 299-300 (1974).
5. Fox, P. J., Pitman, W. C. III & Shephard, F. *Science* **165**, 487-489 (1969).
6. Morgan, W. J. *J. geophys. Res.* **73**, 1959-1982 (1968).
7. Turcotte, D. L. *J. geophys. Res.* **79**, 2573-2577 (1974).
8. Somers, M. L. et al. in *Proc. IEEE/IERE Sub-conference on Offshore Instrumentation and Communications*, Tech. Session J., 16-24 (Oceanology International, London, 1978).
9. Minster, J. B. & Jordan, T. H. *J. geophys. Res.* **83**, 5331-5354 (1978).
10. Collette, B. J., Slootweg, A. P., Verhoef, J. & Roest, W. R. *Proc. K. ned. Akad. Wet.* **B87**, no. 1 (1984).
11. Collette, B. J., Verhoef, J. & de Mulder, A. F. J. *J. geophys.* **47**, 91-98 (1980).
12. Collette, B. J., Schouten, H., Rutten, K. & Slootweg, A. P. *Mar. geophys. Res.* **2**, 143-179 (1974).

The Gloria survey conducted by MV *Farnella* was funded by the UK NERC. The Dutch programmes are funded by the Netherlands Organization for the Advancement of Pure Scientific Research (ZWO) and the Netherlands Council (formerly Commission) for Ocean Research. We acknowledge the assistance of colleagues, students and the crews of all ships involved in the data gathering. A. P. Slootweg computed the fossil flow line pattern.

13. Ramberg, I. B., Gray, D. F. & Reynolds, R. G. *Bull. geol. Soc. Am.* **88**, 609-620 (1977).
14. Collette, B. J. in *26th Congr. géol. int., Abstr. II*, 715, Paris (1980).
15. Lowrie, W. & Alvarez, W. *Geology* **9**, 392-397 (1981).
16. McKenzie, D. P. & Sclater, J. G. *Bull. Volcanol.* **33**, 101-118 (1969).
17. Sclater, J. G. & Francheteau, J. *Geophys. J. R. astr. Soc.* **20**, 509-542 (1970).
18. Parsons, B. & Sclater, J. G. *J. geophys. Res.* **82**, 803-823 (1977).
19. Richter, F. & McKenzie, D. *J. Geophys.* **44**, 441-471 (1978).
20. Lister, C. R. B. *Nature* **257**, 663-665 (1975).
21. Luyendyk, B. P. & MacDonald, K. C. *Bull. geol. Soc. Am.* **88**, 648-663 (1977).
22. Choukroune, P., Francheteau, J. & Le Pichon, X. *Bull. geol. Soc. Am.* **89**, 1013-1029 (1978).
23. Atwater, T. & MacDonald, K. C. *Nature* **270**, 715-719 (1977).
24. Searle, R. C. *Nature* **274**, 187-188 (1978).
25. Collette, B. J. & Slootweg, A. P. *Nature* **274**, 187 (1978).
26. Purdy, G. M. & Rabinowitz, P. D. *Init. Rep. DSDP Leg 45* (1978).

Teleseismic prospecting of lithospheric contrasts beneath the Pyrenees and Alps

Alfred Hirn*, Georges Poupinett†, Gérard Wittlinger‡, Josep Gallart*§ & François Thouvenot†

*Laboratoire d'Etude Géophysique de Structures Profondes, associé au CNRS, Institut de Physique du Globe, 4 place Jussieu, 75230 Paris Cedex 05, France

†Laboratoire de Géophysique Interne, IRIGM, Grenoble, France

‡Laboratoire de Seismologie, Institut de Physique du Globe, Strasbourg, France

§Departamento de Fisica de la Tierra y del Cosmos, Universidad de Barcelona, Barcelona 7, Spain

The possibility of deriving the three-dimensional structure of the lithosphere by inversion of arrival times from distant earthquakes has previously been assessed using data from permanent seismic arrays¹. Temporary emplacement of arrays in regions of key interest furnishes a practicable tool for investigating their structure^{2,3}, but the data needed for resolving complex three-dimensional structures are difficult to collect. We have therefore selected here regions where cylindrical symmetry reduces the problems to two dimensions and where the geological structure of interest is likely to produce large, easily detectable, relative arrival time anomalies. In this respect the depth variation of an interface with a large velocity contrast such as the Moho, as it is known to exist on average perpendicularly to the axis of alpine mountain ranges like the Alps and the Pyrenees, is of interest. In the case of the Pyrenees, explosion seismology profiling on parallel east-west lines in the North Pyrenean Zone (NPZ) and the Palaeozoic Axial Zone (PAZ)⁴⁻⁷ provides evidence of an increase of over 15 km in Moho depth across the range. The variation in this depth beneath the North Pyrenean Fault (NPF)⁸, a feature clearly expressed at the surface in the east as separating the NPZ and PAZ, can be shown to be very sharp, occurring over approximately 5 km. These results call into question the hitherto assumed smoothness of deep structure.

The temporary array of seismic stations (Fig. 1) was maintained from the end of October 1982 to the beginning of January 1983. Ten earthquakes at distances of between 50° and 150° and with magnitudes between 5.1 and 6.3 were recorded by 8 to 18 stations. Arrival times undergo such sharp variations within the array that we cannot define a finite value of the wave slowness which could be used to infer smooth and deep structural variations⁹. Instead the data can be reduced to arrival time delays

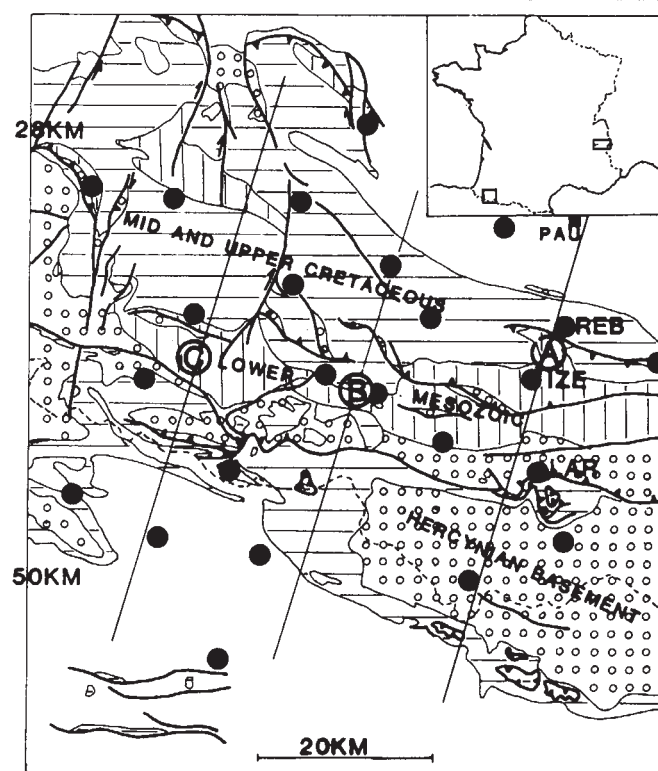


Fig. 1 Temporary seismic stations in the Western Pyrenees. Depths to the crust-mantle interface, from east-west refraction profiles, are 28 km in the north, 50 km in the south. Epicentres of the three largest earthquakes in the region over the past 20 yr are indicated as A, B and C. Relative teleseismic residuals projected on lines through A, B and C are given in Fig. 2.

for each single station with respect to a standard propagation table of seismic waves¹⁰. Most stations were situated either on outcropping crystalline rocks or on basement with a thin cover of high-velocity pre-Cretaceous sediments. The northernmost stations were underlain by several kilometres of Cretaceous flysch and more recent sediments. Sedimentary delays can be accurately computed, reaching values as high as 1.0 s. For an event in Kazakhstan (distance 53°, azimuth 53°) a sharp change in the relative delays of the order of 1 s occurs between the northern (early) and southern (late) stations over a distance of 15 km (station IZE to LAR). For a steeper incidence (Hebrides, distance 148°, azimuth 23°) a similar delay exists but this time IZE falls in the group of late stations. When projected back to the place where the rays cross the Moho (Fig. 2a) a cross-section is obtained which represents the Moho topography assuming

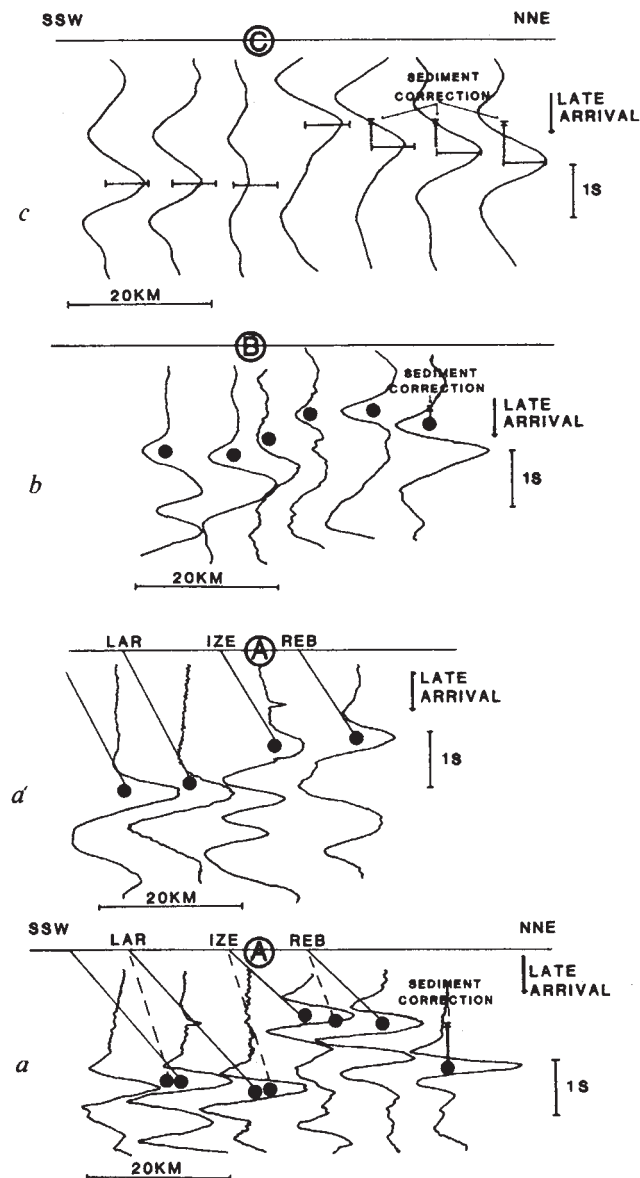


Fig. 2 Relative teleseismic residuals across the Pyrenees. One second of relative residual corresponds to a change of ~ 20 km in Moho depth (vertical dimensions are thus compressed by 2.5 with respect to horizontal). *a*, Section through point A (see Fig. 1): P waves from Kazakhstan event projected back to where they emerge from the mantle. Dots corresponding to New Hebrides earthquake with steeper incidence are added to show the sharpness of spatial variation in structure as data for IZE belong to early arrivals for the first and late arrivals for the second event. Correction for known sediments beneath the northernmost station account for 0.8 s later arrival. Note the change in delay beneath point A, indicating more than 20 km of variation in Moho depth. *a'*, Section through point A: seismograms of Salomon Islands earthquake intermediate in azimuth and incidence between the previous ones, with complex recording at IZE indicative of multipathing. *b*, Section through point B: seismograms are from Ecuador earthquake. The step to early arrivals (thinner crust) for northern stations is less sharp as here waves approach from the WSW where the crust is thicker. *c*, Section through point C: cross-correlograms of emergent waves from Kamchatka earthquake with a reference station. Note the change in delay beneath point C indicating also more than 20 km change in Moho depth. Taking into account known sediments beneath the three northern stations would bring arrivals as early as for the fourth one, indicating an almost flat shallow Moho.

the relative delays are wholly caused by its variation in depth. The 1-s difference between the southern and northern stations is consistent with the 20–25 km difference in Moho depth derived from explosion seismology^{4–8}, if the slightly larger

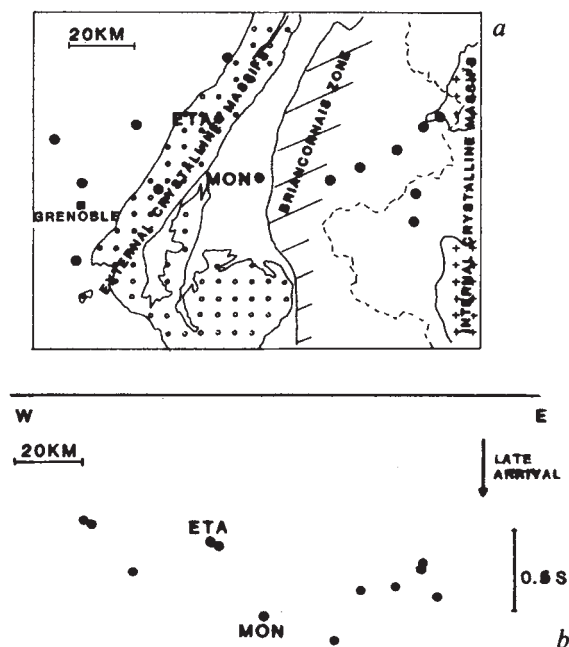


Fig. 3 *a*, Temporary seismic array in the Western Alps. *b*, Relative residuals of waves arriving due north from Aleutian earthquakes on a west-east section of the array. Note the relative delay from ETA to MON.

average crustal velocity in the north is taken into account. Our data also indicate that this variation occurs very sharply on a north-south distance of the order of 10 km, near point A. This is additionally documented by a Salomon Islands earthquake, intermediate in azimuth and incidence between the previous two which at station IZE gives a complex seismogram indicative of multipathing (Fig. 2*a'*). This is evidence for a situation similar to that in the eastern part of the range—juxtaposition of two crustal segments of different thickness—and it illustrates the westward extension of the deep expression of the NPF. In Fig. 2*b* seismograms from an earthquake in Ecuador recorded by central stations in the array are displayed on a line through point B. Waves arrive from the WSW, almost opposite to the direction of approach from Kazakhstan. The step to earlier arrivals for the northern stations is also documented here although it appears less clearly because the waves are approaching through thicker crust. Figure 2*c* presents a picture similar to Fig. 2*a* for the western group of stations in the array from an earthquake in Kamchatka (distance 82° , azimuth 12°); as the signals have progressive onsets and a longer period, the cross-correlograms have been plotted, projected back to where the rays cross the Moho. Beneath point C and within 10 km horizontally, a 1 s step is again documented, enabling extension of the deep expression of the NPF even further west. Tracing the NPF as a deep fracture further than seen at the surface provides additional support for evolutionary models emphasizing its fundamental regional role^{11,12}. Extension of the NPF to the western half of the range is a subject of debate among geologists as surface evidence is not clear here although seismic activity is well documented¹³.

As to the present state of the NPF, points A and B and C where we document the Moho disruptions were the exact epicentres of the three largest earthquakes which occurred in the western half of the Pyrenees over the past two decades—those at Arudy, 29 February 1980, Mb ~ 5.1 (ref. 14); at Arette, 13 August 1967, Mb ~ 5.5 ; and at Arbailles, 6 January 1982, Mb ~ 4.8 . We therefore propose that the location of the main seismic activity in the present-day stress-field is controlled by preexisting heterogeneity in the inherited lithospheric structure.

Reconnaissance explosion seismology in the western Alps in the 'fifties and early 'sixties led to contour maps of Moho depth^{15,16} which in this region showed an important but smooth

dip of the Moho of 20% from west to east; however, the data constraining these contours are generally loose. All models of the deep structure of the western Alps have this smooth Moho with variable upper crustal structures thought to correspond to complicated overthrusts, with westward flaking of crustal segments, matching the surface decollements and nappes. Such models are used not only for the shallow mantle zone of Ivrea, but also for even further west than the external crystalline massifs¹⁷. We maintained our array for 6 weeks as an essentially west-east line from the foreland through the external crystalline massifs and the Briançonnais unit (Fig. 3a).

The most unexpected result of our teleseismic experiment for waves arriving from the north was a clear delay across the boundary between the crystalline massifs of the external zone and the Briançonnais unit of the Penninic zone to the east (Fig. 3b). The length of this delay (0.6 s) and the sharpness of the variation, which occurs between two stations 15 km apart, are similar to those across the NPF. Thus our data suggest that in the western Alps the deep structure may in places be similar to that of the Pyrenees, more closely resembling a juxtaposition of lithospheric segments than a smooth transition between them. Part of the anomaly might be explained by supposing that the crusts on either side differ in average velocity and composition, as has been proposed further to the west¹⁷, or that mantle velocities are different. Nevertheless it seems difficult to account for the observations without supposing a step in an interface with a strong velocity contrast such as the Moho, which we consider better explains both the amount and sharpness of the anomaly.

Brief periods of recording a small number of teleseismic signals by a dense array in two geologically interesting regions of supposed lateral variation in crustal thickness have shown

sharp and strong delay time anomalies. In the Pyrenees a problem is solved: the NPF remains a major fracture, with a step in the Moho in at least part of the western half of the range where its surface expression is not clear and where explosion seismology has not allowed precise definition. The coincidence of the surface projection of the Moho step, as detected from teleseismic anomalies, and the epicentres of the main earthquakes at upper and middle crustal depths imply that the lithospheric fracture extends almost vertically to the surface. In the case of the Alps, detection of similar variations in delay times brings into question the hitherto assumed smoothness of Moho depth variation and prompts the setting up of further specially designed explosion seismology experiments as used in the Pyrenees⁴⁻⁷.

Seismic stations were designed and maintained by René and Jean Verhille and Daniel Brefort. This paper is contribution IPG no. 718.

Received 6 July 1983; accepted 16 January 1984.

1. Aki, K., Christofferson, A. & Husebye, E. S. *J. geophys. Res.* **82**, 277-296 (1977).
2. Briden, J. C. *et al. Nature* **292**, 123-128 (1981).
3. Dörbath, C., Dörbath, L., Le Page, A. & Gaulon, R. *Ann. Géophys.* **1**, 25-36 (1983).
4. Explosion Seismology Group Pyrenees *Ann. Géophys.* **36**, 135-140 (1980).
5. Gallart, J., Daignières, M., Banda, E., Surinach, E. & Hirn, A. *Ann. Géophys.* **36**, 141-158 (1980).
6. Daignières, M., Gallart, J. & Banda, E. *Ann. Géophys.* **37**, 435-456 (1981).
7. Gallart, J., Banda, E. & Daignières, M. *Ann. Géophys.* **37**, 457-480 (1981).
8. Hirn, A., Daignières, M., Gallart, J. & Vadell, M. *Geophys. Res. Lett.* **7**, 263-266 (1980).
9. Briden, J. C., Mereu, R. F. & Whitcombe, D. N. *Geophys. J. R. astr. Soc.* **71**, 793-808 (1982).
10. Herrin, E. *Bull. seism. Soc. Am.* **58**, 1193-1241 (1968).
11. Choukroune, P. & Mattauer, M. *Bull. Soc. Géol. Fr.* **20**, 689-700 (1978).
12. Daignières, M., Gallart, J., Banda, E. & Hirn, A. *Earth planet. Sci. Lett.* **57**, 88-100 (1982).
13. Gagnepain, J. *et al. Ann. Géophys.* **36**, 499-508 (1980).
14. Gagnepain-Beyneix, J., Haessler, H. & Modiano, T. *Tectonophysics* **85**, 273-290 (1982).
15. Perrier, G. thesis, Univ. Paris (1973).
16. Giese, P., Morelli, C. & Steinmetz, L. *Tectonophysics* **20**, 367-379 (1973).
17. Thouvenot, F. & Perrier, G. *Pure appl. Geophys.* **119**, 163-184 (1980/1981).

Time-dependent models of single- and double-layer mantle convection

Alan Paul Boss & I. Selwyn Sacks

Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington, DC 20015, USA

One outstanding problem of great geophysical importance is the vertical extent of convection in the Earth's mantle¹. Steady-state models of convection in the mantle have yet to produce a description of the mantle which is consistent with geochemical and geophysical observations. Here we present time-dependent numerical models of two-dimensional convection which simulate mantle convection. Starting from a fluid initially at rest with a purely conductive temperature profile, we now find that for a Rayleigh number (Ra) of 10^7 , the model experiences a transient period of double-layer convection, which lasts on the order of hundreds of millions of years when scaled to the Earth's mantle. It is suggested that transient periods of double-layer convection in the Earth's mantle may have existed in the past, whether or not such a period exists today.

Models of mantle convection have been proposed with several different combinations of convecting layers, including convection confined to the upper mantle, convection in separate layers in the upper and lower mantle, and whole-mantle convection²⁻¹⁰. The models of double-layer convection have been motivated by geochemical analyses which suggest the presence of two (or more) sources of elementally and isotopically distinct rock, possibly the mantle above and below the 670-km seismic discontinuity¹¹⁻¹⁴. This discontinuity may be caused by a phase transition or by a change in the chemical composition. Phase changes alone are unlikely to prevent convection across the 670-km discontinuity^{15,16}, but only a small increase in chemical

density could prevent whole-mantle convection^{16,17}, though some mixing between the layers would still occur (P. Olson, personal communication). However, steady-state, two-layer convection implies a substantial temperature drop across the intervening conductive boundary layer, which requires a much lower viscosity in the lower mantle than in the upper mantle⁹. The latter is inconsistent with results which show the viscosity of the mantle to be nearly uniform¹⁸.

Convection is in general a time-dependent phenomenon, rather than steady-state. Previous models of time-dependent convection have produced only a single layer of convecting cells¹⁹⁻²⁵; often the time dependence consisted of an oscillation about a steady-state pattern. However, our models²⁶ show that as Ra is increased towards values appropriate for the present mantle, qualitatively different evolutions result, at least for certain initial conditions.

The models are based on the Boussinesq approximation at infinite Prandtl number, and treat convection in a two-dimensional, cartesian coordinate box (aspect ratio of 1.4), heated from below. The equations of motion are solved numerically by a computer code originally written by Lux²⁷. The biharmonic equation is solved directly by the method of conjugate gradients, while the energy equation is solved by an alternating directions-implicit method with upwind differences for the advective terms. The viscosity is taken to be exponentially dependent on the temperature (proportional to $\exp[6(0.5 - T)]$), in terms of the dimensionless temperature T ²⁸, so that a variation by a factor of 400 is possible between the hottest ($T = 1$) and coldest ($T = 0$) regions. This variation is intended to simulate at least partially the strong temperature and pressure dependence of the viscosity presumed to occur for mantle matter¹. The Rayleigh number for these variable viscosity models is based on the viscosity for $T = 0.5$, which occurs at the centre of the layer in steady state. The boundary conditions are free slip, with impenetrable, insulating side walls. The time step is at most a fraction of the Courant-Friedrichs-Lewy time step, ensuring temporal resolution of time-dependent flows. The spatial resolution of the models is sufficient to include at least three grid points in the

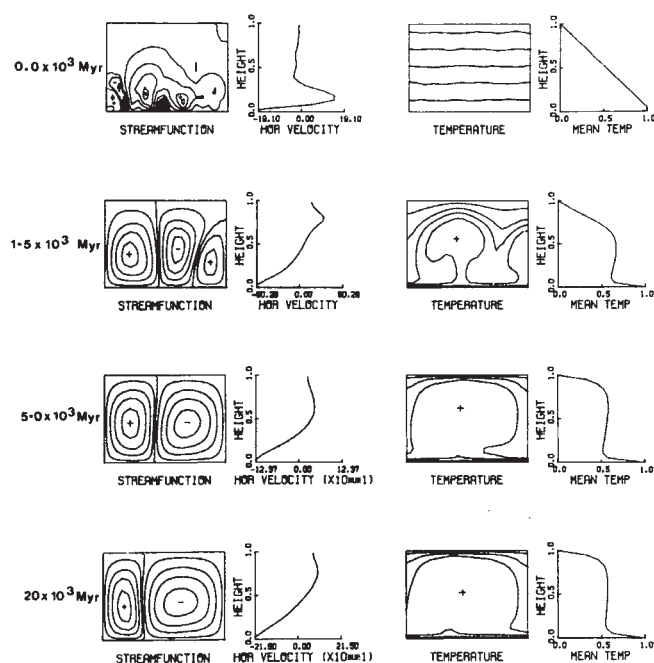


Fig. 1 Time evolution of a model of mantle convection with $Ra = 10^5$. The contours of the streamfunction are shown, together with the mean horizontal velocity as a function of depth, isotherms of the temperature distribution, and the mean temperature as a function of depth. The streamfunction is contoured at values of 0 and $1/8, 3/8, 5/8$ and $7/8$ times the minimum and maximum, and the isotherms at temperatures of $1/10, 3/10, 5/10, 7/10$ and $9/10$ times the maximum. +, Local maxima; -, local minima. The horizontal velocity may be converted to a dimensioned velocity by multiplying by $10^{-3} \text{ cm yr}^{-1}$.

boundary layers at the top and bottom of the box, which should result in a reasonably accurate solution²¹ (see also Fig. 3 of Schubert *et al.*²⁹). The code has been tested by comparison with previous calculations of both steady-state^{21,28} and time-dependent convective patterns²¹.

The two models start from a state of pure conduction, that is, a uniformly increasing temperature gradient from the top ($T = 0$ in dimensionless units) to the bottom ($T = 1$) of the box. Similar initial conditions have been used by previous authors^{21,29}. Heating due to impacts during accumulation of the Earth probably resulted in a molten or partially molten planet, with simultaneous formation of the iron core and downward advection of thermal energy³⁰. Convection in partially molten regions would be very efficient and would rapidly reduce the temperature to the solidus, which might then be considered as an appropriate initial temperature for sub-solidus convection. Because the solidus is thought to rise faster with depth than the adiabat³¹, in the Boussinesq approximation the initial condition for T would be a roughly linear increase with depth ($T_i = T_{\text{solidus}} - T_{\text{adiabat}}$), as we have assumed. The streamfunction was given a very small initial value (10^{-8} in dimensionless units—essentially static) and a random pattern of noise (at the 5% level) was superimposed on the initial temperature and streamfunction.

Figure 1 shows the time evolution of the first model, with $Ra = 10^5$. The initial conditions for the two models ($0.0 \times 10^3 \text{ Myr}$) are so far from steady state that initially a burst of convection occurs, characterized by the release of a large plume of hot fluid, as shown in Fig. 1 at $1.5 \times 10^3 \text{ Myr}$. The plume rises to the top of the cell and spreads out, leaving behind a single layer of convective cells in the nearly isothermal core of the box. The convective pattern only slowly approaches the presumed steady-state configuration of a single cell²⁷; even at $20 \times 10^3 \text{ Myr}$ two cells are competing with each other. The time evolution of the averaged temperature as a function of depth

is similar in this model to that obtained by Schubert *et al.*²⁹ for Ra of $\sim 10^4$ – 10^5 .

The second model is identical to the first model, except that now $Ra = 10^7$ and the initial noise differs because of the increased number of grid points (51×51 as opposed to 19×19). This value of Ra is close to that appropriate for whole-mantle convection with a viscosity of 10^{22} P (ref. 18). Figure 2 shows that the initial conditions again lead to the formation of a hot plume which rises to the top of the layer (15 Myr ; see Fig. 1 at $1.5 \times 10^3 \text{ Myr}$). Thereafter, the evolution diverges greatly from the lower Ra model. A chaotic sequence of double-layer convective cells results (hinted at 47 Myr , clearly delineated at 510 Myr). Eventually the double-layer pattern disappears ($1.2 \times 10^3 \text{ Myr}$), leaving behind a single layer of cells ($1.9 \times 10^3 \text{ Myr}$).

The physical reason for the double layers of convection may be found by considering the temperature as a function of depth. Multiple layering occurs in regions with inverted temperature gradients where the temperature decreases with depth (in vertical sections, as well as in the mean). This occurs in the second model because with the chosen initial conditions, the initial burst of convection overshoots the steady-state temperature profile, and results in a large amount of hot fluid reaching the top of the layer. This initial burst is accomplished in a single layer of convection cells, as that is the most efficient way in which to remove the initial excess of hot fluid at the bottom of the layer. Once the inverted temperature gradient is obtained, the intermediate layer of the box is stable with respect to convection.

The stability of the inverted temperature gradient region may be demonstrated by considering the buoyancy of fluid elements. In the Boussinesq approximation, the buoyancy (acceleration) of a fluid element is dependent only on its temperature. Consider first the lower boundary layer. The effective buoyancy of the lower boundary layer will be proportional to the average temperature (T_b) of the layer. Because of the shape of the boundary layer, T_b will have a value somewhat less than that of the temperature halfway between the lower boundary value and the minimum in the inverted region. Compare this temperature with the maximum temperature (T_i) in the inverted layer. If the fluid in the bottom boundary layer is to be able to penetrate to the top of the box, its buoyancy must exceed that of all the overlying layers, that is, T_b must exceed T_i . In the model with $Ra = 10^5$, T_b exceeds T_i at all times, and hence a single layer of convection results. However, in the model with $Ra = 10^7$, T_b is less than T_i by 47 Myr , so that until this inequality is reversed, the inverted layer is more buoyant than the bottom boundary layer, and a single layer of cells is prohibited. This argument is readily extended to the upper boundary layer (only in terms of negative buoyancy).

In the model with $Ra = 10^7$, the temperature gradients at the top and bottom of the box continue to drive convective cells of limited vertical extent. Note that for $Ra = 10^7$, a layer of only one-tenth the total depth is still convectively unstable ($Ra > Ra_c$). The inverted temperature gradient is gradually eroded by these convective motions. The convective cells in the top and bottom layers grow into the previously stable layer, and eventually merge into a single layer of cells. The transient phase of double-layer convection lasts on the order of $0.5 \times 10^3 \text{ Myr}$. At the higher values of Ra appropriate for the early Earth, it is uncertain whether this time would be increased (more unstable to overshoot) or decreased (more vigorous convection for removing the inversion).

The inverted temperature gradient serves as a barrier to whole-mantle convection, and thereby helps prevent the mixing that would homogenize any isotopic or elemental differences that might arise between the layers. The multiple layers occur because of the presence of an intermediate layer having an inverted temperature gradient, which is stable with respect to fluid elements in the boundary layers. The magnitude of the inverted gradient increases as Ra increases, so that at a new, critical value of Ra , convection is unable to penetrate the stable layer. We have shown that this new critical value lies between 10^5 and 10^7 ; further calculations will be necessary to define this

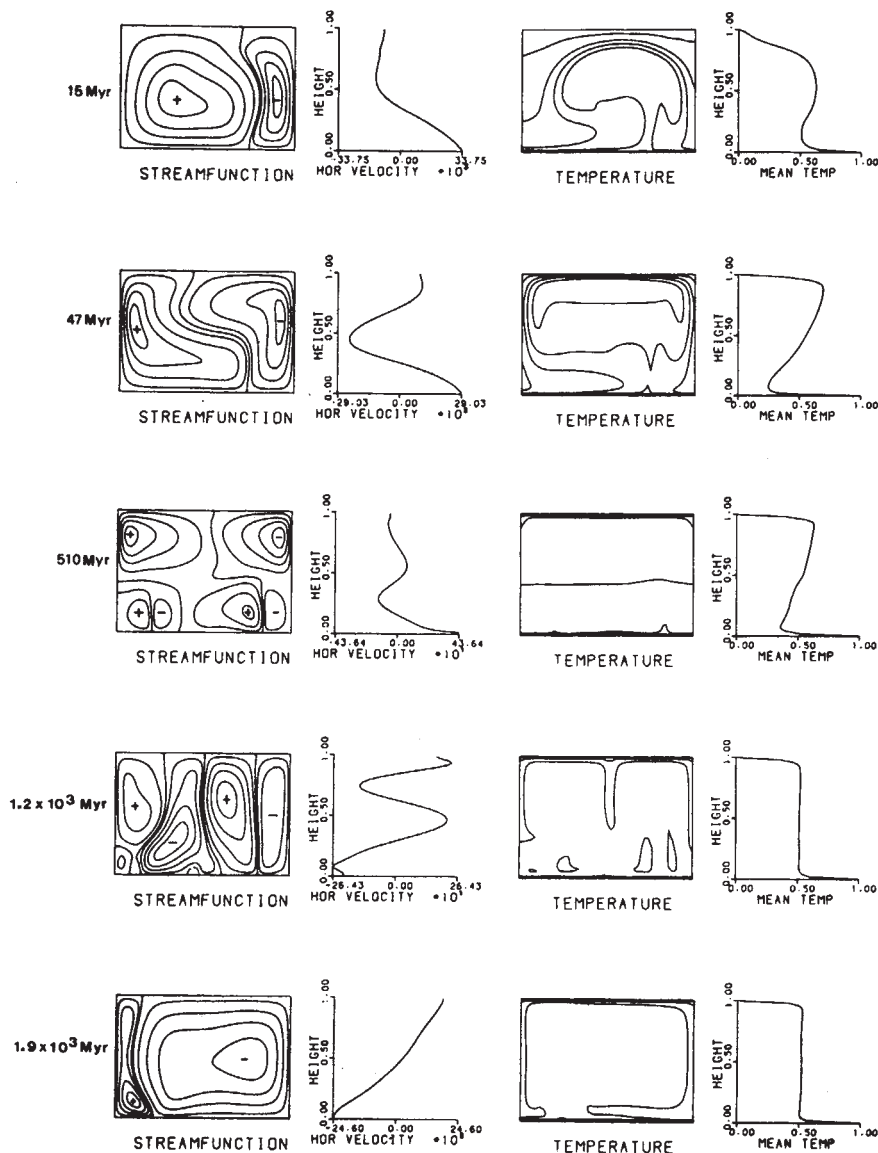


Fig. 2 Time evolution of a model of mantle convection with $Ra = 10^7$ (see Fig. 1 legend). Note the presence of a transient period of double-layer convection, evidenced by the streamfunction (closed streamlines of the fluid flow) at 510 Myr.

value better. Thus, this mechanism provides another means, besides chemical density changes, of enforcing stratification of convective layers, though only in a transient manner. While these results are possibly significant for models of the Earth's evolution; they should be verified on higher spatial resolution grids, as well as with different numerical methods. The assumed initial conditions are critical for inversion and thus require further investigation.

This work was partially supported by NASA.

Received 21 November 1983; accepted 14 February 1984.

- Schubert, G. A. *Rev. Earth planet. Sci.* **7**, 289-342 (1979).
- Davies, G. F. *Geophys. J. R. astr. Soc.* **49**, 459-486 (1977).
- Elsasser, W. M., Olson, P. & Marsh, B. D. *J. geophys. Res.* **84**, 147-155 (1979).
- Jeanloz, R. & Richter, F. M. *J. geophys. Res.* **84**, 5497-5504 (1979).
- Olson, P. *J. geophys. Res.* **86**, 4881-4890 (1981).
- Richter, F. M. & McKenzie, D. P. *J. geophys. Res.* **86**, 6133-6142 (1981).
- Schubert, G. & Spohn, T. *Geophys. Res. Lett.* **8**, 951-954 (1981).
- Fowler, A. C. *Geophys. Res. Lett.* **9**, 816-819 (1982).
- Kenyon, P. M. & Turcotte, D. L. *J. geophys. Res.* (submitted).
- Spohn, T. & Schubert, G. *J. geophys. Res.* **87**, 4682-4696 (1982).
- Hofmann, A. W., White, W. M. & Whitford, D. J. *Carnegie Instn Wash. Yb.* **77**, 548-562 (1978).
- DePaolo, D. J. *Nature* **291**, 193-196 (1981).
- Jacobsen, S. B. & Wasserburg, G. J. *Tectonophysics* **75**, 163-179 (1981).
- Allegre, C. J. *Tectonophysics* **81**, 109-132 (1982).
- Christensen, U. *Geophys. J. R. astr. Soc.* **68**, 487-497 (1982).
- Olson, P. & Yuen, D. A. *J. geophys. Res.* **87**, 3993-4002 (1982).
- Richter, F. M. & McKenzie, D. P. *J. geophys. Res.* **86**, 6133-6142 (1981).
- Peltier, W. R. A. *Rev. Earth planet. Sci.* **9**, 199-225 (1981).
- Andrews, D. J. *J. geophys. Res.* **77**, 6470-6481 (1972).
- Young, R. E. *J. Fluid Mech.* **63**, 695-721 (1974).
- McKenzie, D. P., Roberts, J. M. & Weiss, N. O. *J. Fluid Mech.* **62**, 465-538 (1974).
- Schubert, G. & Young, R. E. *Tectonophysics* **35**, 201-214 (1976).
- Daly, S. F. *Geophys. J. R. astr. Soc.* **61**, 519-547 (1980).
- Arkani-Hamed, J., Toksoz, M. N. & Hsu, A. T. *Tectonophysics* **75**, 19-30 (1981).

- Houseman, G. & McKenzie, D. P. *Geophys. J. R. astr. Soc.* **68**, 133-164 (1982).
- Boss, A. P. & Sacks, I. S. *Carnegie Instn Wash. Yb.* **81**, 543-550 (1982).
- Lux, R. A. & Sacks, I. S. *Carnegie Instn Wash. Yb.* **79**, 512-520 (1980).
- Torrance, K. E. & Turcotte, D. L. *J. Fluid Mech.* **47**, 113-125 (1971).
- Schubert, G., Cassen, P. & Young, R. E. *Phys. Earth planet. Interiors* **20**, 194-208 (1979).
- Stevenson, D. J. *Science* **214**, 611-619 (1981).
- Basaltic Volcanism Study Project. *Basaltic Volcanism on the Terrestrial Planets*, 1163 (LPI, Houston, Texas, 1981).

Stagnant layers at the bottom of convecting magma chambers

Claude Jaupart, Geneviève Brandeis & Claude J. Allègre

Laboratoire de Géochimie et Cosmochimie, Institut de Physique du Globe et Département des Sciences de la Terre, Université Paris 6 et 7, 4 place Jussieu, 75230 Paris Cedex 05, France

The evolution and crystallization of igneous complexes has received much attention from petrologists¹⁻³ and more recently from physicists⁴⁻¹⁰. Current models emphasize the role of compositional effects. We present here a different viewpoint. Because compositional effects are due to crystallization, they depend on the thermal structure and regime of cold boundary layers in convecting magma chambers, particularly of the bottom layer where the thickest rock sequences form. We have studied purely thermal convection in a large aspect ratio magma chamber which is cooled through both its upper and lower boundaries.

We made laboratory fluid dynamical experiments in turbulent and transient conditions and show that a stagnant layer develops at the bottom, isolated from the convective part of the chamber. The essential features of this layer are that it is not affected by mixing and that a significant thermal gradient is maintained across it. These imply peculiar crystallization conditions.

We have studied the sudden cooling of an initially hot and isothermal layer of fluid. The fluid is placed in a Plexiglass tank of 25×25 cm horizontal dimensions and 10 cm height. The tank walls are 2 cm thick to minimize lateral heat losses. Fixed temperature boundary conditions are achieved using two 3-cm thick copper plates at the upper and lower surfaces, through which thermostated water circulates. At the beginning of the experiment, temperature is uniform in the fluid and in both copper plates at a value T_0 . At time $t = 0$, the copper plates are switched to a cold water bath. They are cooled rapidly, reaching 90% of their final temperature in about 2.5 min. Temperature is maintained constant at a lower value $T_0 - \Delta T$ to within 0.1°C after 6 min. These conditions approximate those of instantaneous cooling.

In the standard Boussinesq approximation and for the fixed temperature boundary conditions of interest here, two non-dimensional parameters characterize the experiments^{11,12}: Rayleigh number $Ra = (g\alpha\Delta Td^3)/\kappa\nu$ and Prandtl number $Pr = \nu/\kappa$, where d is the thickness of the layer, g the acceleration due to gravity, α the coefficient of thermal expansion, κ thermal diffusivity and ν kinematic viscosity. We assume that viscosity is constant (in our experiments, it never varies by more than 30%). In natural systems, viscosity depends strongly on temperature. For such fluids, theoretical calculations as well as laboratory experiments have shown that convective flows are limited to regions where viscosity variations are small and resemble the uniform viscosity flows¹³⁻¹⁵. We have investigated the constant viscosity problem as it is a prerequisite to a proper understanding of the more complex variable viscosity problem.

We do not include compositional effects which result from crystallization and therefore depend on the cooling regime.

The physical properties of basaltic liquids¹⁶⁻¹⁹ indicate that representative Prandtl numbers exceed 4,000. For a temperature difference of 50°C and a chamber height of 1 km, the Rayleigh number is extremely large (10^{16}). We used silicone oil with kinematic viscosities ranging over 10^2 – 10^3 mm² s⁻¹. For those, Pr ranges from 800 to 8,000. The proper scaling of our experiments can be made using the analysis of Kraichnan for high Prandtl number fluids¹². The range of Rayleigh numbers investigated is 10^6 – 10^8 and molecular viscosity dominates the whole flow structure. This is always the case in boundary layers, whatever the Rayleigh number value, and it is this that we have investigated.

Several types of observations were made. Shadowgraph pictures allow the visualization of the whole temperature structure (see ref. 17). To obtain a more detailed picture of convection, we also used differential interferometry²⁰ with a 4-cm-wide laser beam. This gives a picture of the iso-gradient temperature lines to an accuracy of about $0.01^\circ\text{C cm}^{-1}$. Finally, we placed a series of 11 thin (0.2 mm diameter) platinum wires stretched horizontally across the fluid. These allow the measurement of the mean horizontal temperature to an accuracy of about 0.2°C , or typically 1% of the overall temperature difference.

We have run more than 15 experiments with different oils and varying temperature differences. All showed the same effects, which we now describe for one experiment run with $\Delta T = 22.2^\circ\text{C}$ and $Ra = 5.9 \times 10^6$. Convective instability starts at about 2 min (Fig. 1a), when a local Rayleigh number based on the thermal boundary layer thickness exceeds a value of about 1,600 in agreement with other studies²¹. The boundary layer breakdown occurs in the form of thermals or, more accurately, 'starting plumes'²². These take about 1 min to reach the lower parts of the fluid (Fig. 1b). In this initial stage, they almost penetrate to the lower surface. With time the thermals become

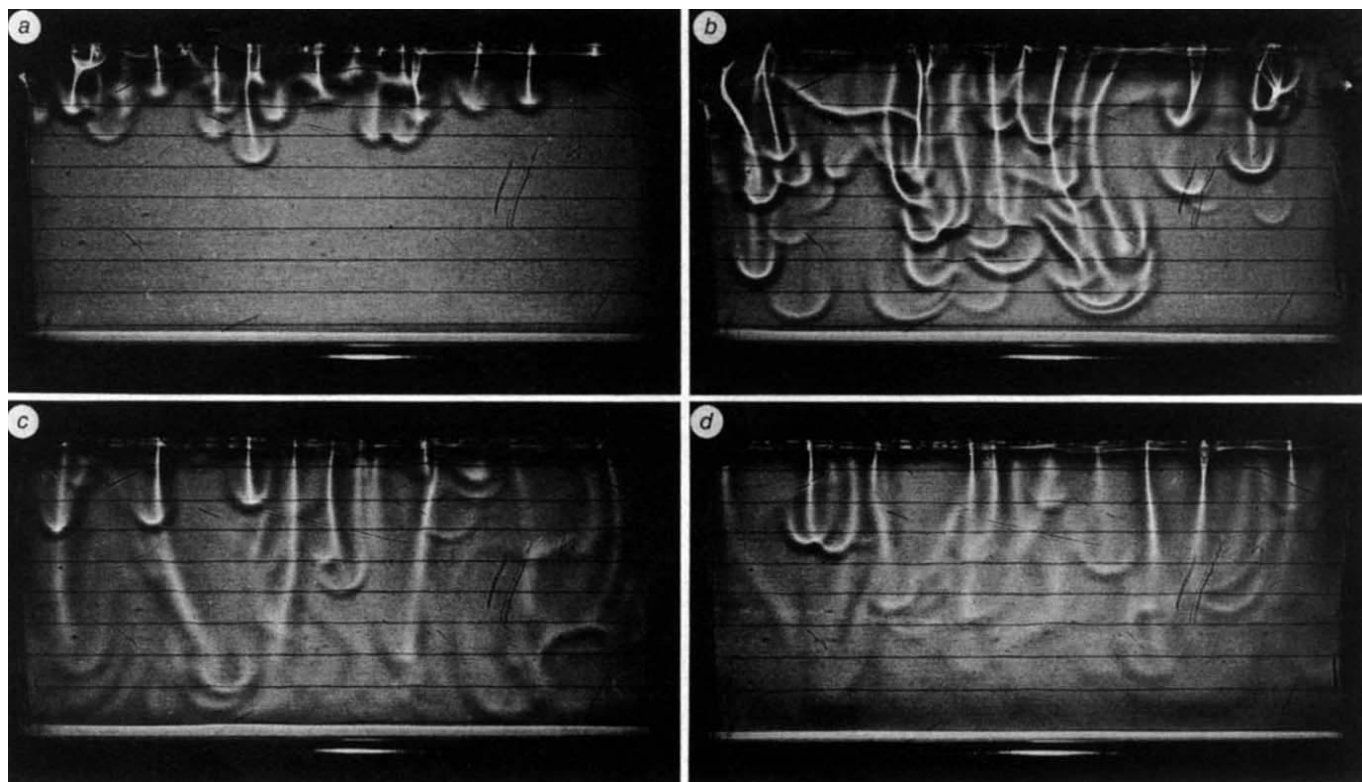


Fig. 1 Time-sequence of the development of convection for the experiment discussed in the text. The lower boundary layer is barely detectable because of light refraction (the wire which is seen at the very bottom lies at 1 cm above the lower surface). The right end of the upper boundary layer goes unstable slightly later than the left end because cooling is not perfectly uniform initially. The effect is not large, as witnessed by the small difference in onset time. *a*, Convection starts: thermals are generated in the upper boundary layer. *b*, The convective thermals or 'starting plumes' reach the bottom. *c*, Fully developed convection (time 4'30"). *d*, Convection in a late stage (time 13'30"). Compare with *c* and *b*. The grey area at the very bottom is not penetrated by thermals: this is the stagnant layer.

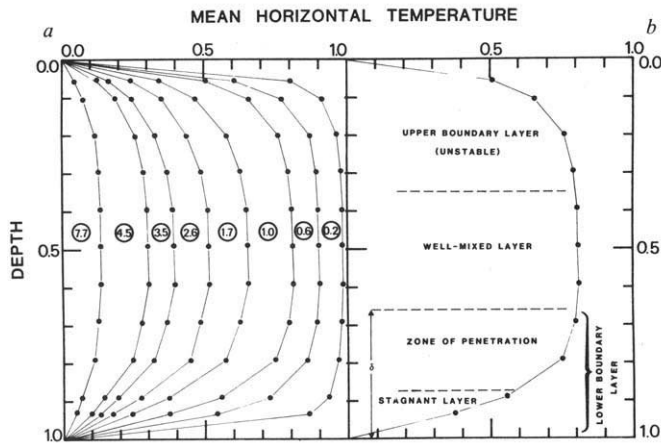


Fig. 2 *a*, Vertical profiles of the mean horizontal temperature as time increases. Numbers along the curves indicate the non-dimensional time, defined as $(d^2 t / \kappa)$, in units of 10^{-2} . Real times are, in minutes: 3, 9, 15, 27, 40, 55 and 90. Steady-state conditions are reached in about 4 h for this experiment. *b*, Schematic structure of convection for the profile at 1.0 (15 min) from both the temperature data and the photographs.

slower and do not reach the bottom (compare Figs 1*b*, *d*): a stagnant layer of significant thickness has formed. The thermals travel in the fluid with an approximately constant velocity until they are strongly decelerated and then stop at the top of the stagnant layer. The zone of deceleration corresponds to that of penetration of the stable temperature gradient at the base of the tank.

The general structure of convection is described in Fig. 2. At the top there is an unstable boundary layer where thermals are generated. In the middle parts of the fluid the mean horizontal temperature is approximately constant (within experimental accuracy). This defines a well-mixed layer. Below this, the lower thermal boundary layer is made of two parts: a zone of penetration where thermals decelerate and a stagnant layer where there is no convection. Note that the temperature gradient is large in the stagnant layer. This is associated with a density increase which does not allow penetration.

The profiles of Fig. 2 indicate that the thickness of the well-mixed layer decreases slowly with time as cooling proceeds and convection becomes more sluggish. This can be shown by the growth of the lower boundary layer. Its thickness was determined by fitting a second-degree polynomial to the temperature

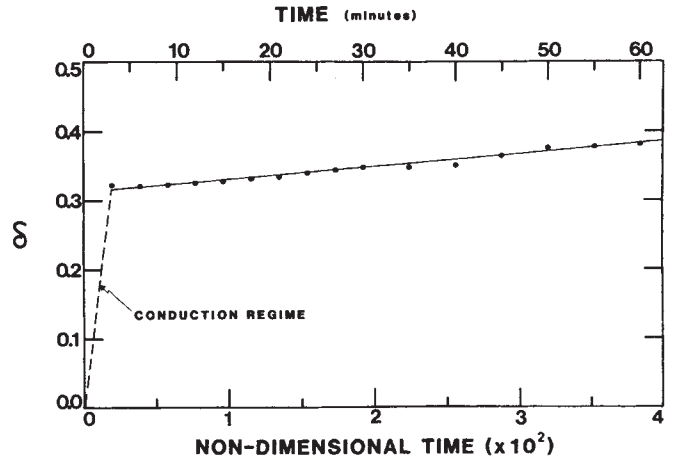


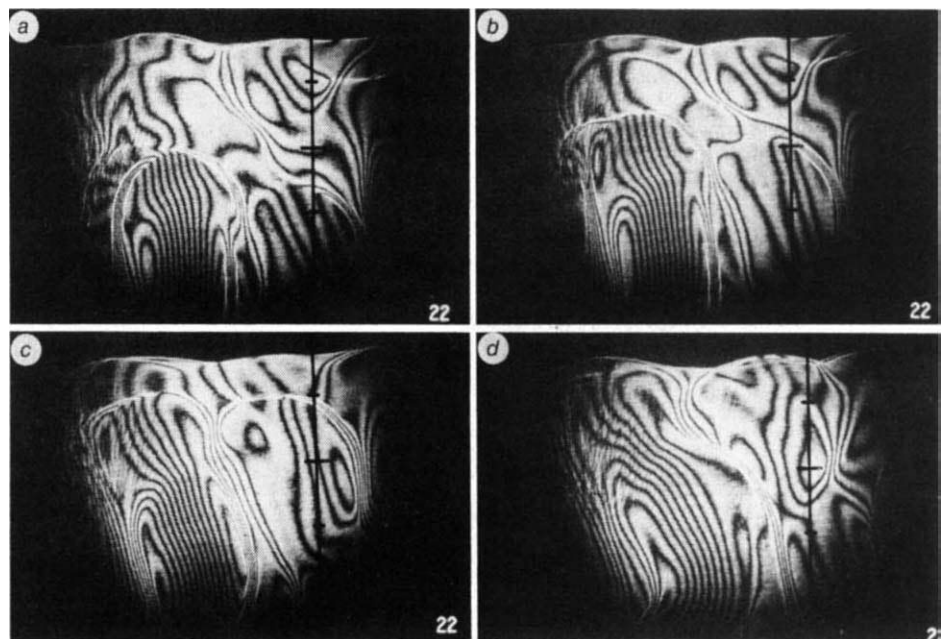
Fig. 3 Evolution of δ , the lower boundary layer thickness (in units of d , the total layer thickness). Note the initial period of fast growth when there is no convection. Within experimental error, the data can be fitted with a straight line.

data and then calculating the height at which temperature has a zero vertical gradient. Figure 3 shows that this thickness evolves very slowly with time after an initial period of formation, which is simply the conduction stage before the convective thermals reach the bottom. The evolution of the stagnant layer is similar. After an initial episode of fast growth, its thickness stabilizes and increases almost imperceptibly.

To confirm the existence of the stagnant layer, we used differential interferometry. Because in the case of cooling the light rays are refracted into the stagnant layer, no clear and simple picture can be obtained. We therefore performed the reverse experiment and suddenly heated an initially cold fluid layer. A stagnant layer was observed at the top, which is shown in Fig. 4 as a convective thermal approaches. Note that it does not penetrate the layer and is dissipated by conduction and by other incoming thermals. In the cooling experiment, the lower boundary layer is colder and therefore more viscous than the rest of the fluid, which impedes the penetration of thermals. However, the formation of the stagnant layer is not a result of viscosity increase, as evidenced by the heating experiment.

The existence of a stagnant layer is a universal feature of all our experiments. This layer grows by the addition of cold fluid

Fig. 4 Differential interferometry for the reverse experiment: a cold layer is heated suddenly from both upper and lower boundaries. ΔT was 10.5°C and Ra 8.5×10^6 . The photographs were taken after 45 min when the temperature difference between the well-mixed layer and both copper plates was 3.9°C . Photographs are taken at 20-s intervals. The marks on the right-hand side are 1 cm apart. The dark region at the top with the wavy interface is the stagnant layer. There is an incoming thermal in *a*. Note the slight distortion of the interface ahead of the thermal cap. The patterns are complex because the laser beam crosses the whole fluid and averages the effect of the thermal with that of the background.



coming from the top. The basic physical principle is simple and has been illustrated in other types of experiments^{23,24}: cold fluid is stable and nothing will make it rise. Cooling through side-walls would enhance the effect. Our results are therefore of general applicability and do not depend on our particular set of initial conditions (for example, whether or not the temperatures of the upper and lower surfaces are lowered by the same amount). The thickness of the stagnant layer may change, but not the fact that it exists. What is more important and new is that the stagnant layer stabilizes rapidly at some thickness. This shows that penetration is also stabilized. The convective system thus evolves towards a stable spatial configuration which then changes slowly. The physics of this are complex, involving transient convection and penetration in a non-uniform environment. The physical and mathematical treatment of the temperature data allow some further developments and scaling laws. These will be made in a more extensive study.

There is some observational support for the existence of a stagnant layer. Donaldson²⁵ suggested that crystals formed in a quiet environment in the Rhum intrusion. Similarly, Jackson¹ argued that textures and structures in the Ultramafic Zone of the Stillwater Complex indicate that magma was stagnant. Our experiments imply peculiar crystallization conditions at the bottom of magma chambers. There, crystals would not be affected by convective mixing phenomena. Furthermore, crystallization would occur with a rather high heat flux imposed from the top (Fig. 2). This represents highly unstable conditions for crystal nucleation and growth and probably leads to cyclic crystallization sequences on the scale of a few centimetres^{26,27}. More generally, the conditions clearly differ from those in a large reservoir as the stagnant layer is of smaller thickness. In particular, crystallization is slower because country rocks must evacuate both the latent heat of crystallization and the imposed heat flux.

The main limitation of our experiments is the absence of compositional effects. Such effects develop in crystallizing regions where there is a high density of crystals. Because magmas are viscous, the associated density instabilities are not instantaneous. What must be ascertained is whether these instabilities develop before or after the formation of the stagnant layer. If they develop after, they would lead to the destruction of the stagnant layer, which would then form again. This process could repeat itself in a quasi-periodic manner.

Our aim has been to investigate one of the simplest fluid dynamical experiments of relevance to the study of magma chamber processes. In conclusion, the stagnant layer isolates the crystallizing region below from the effects of convective mixing. It represents a smaller volume of magma in which compositional gradients may be set up and destroyed. This offers a physical alternative to the reinjection hypothesis to explain relatively rapid cyclic chemical evolution within large magma chambers.

This study was supported by PIRPSEV and by INAG (ATP Geodynamique 2). We thank P. Bergé for advice, M. Girard, L. P. Ricard and A. Lecomte for help in the laboratory, and Y. Bottinga, H. Huppert, B. D. Marsh, S. A. Morse and H. S. Yoder for thoughtful comments. This is IGP Contribution No. 724.

Received 6 October 1983; accepted 5 January 1984.

1. Jackson, E. D. *U.S. Geol. Surv. Prof. Pap.* **358** (1961).
2. Wager, L. R. & Brown, G. M. *Layered Igneous Rocks* (Oliver & Boyd, Edinburgh, 1967).
3. Morse, S. A. *J. Petrol.* **20**, 555–624 (1979).
4. McBirney, A. R. & Noyes, R. M. *J. Petrol.* **20**, 487–554 (1979).
5. Chen, C. F. & Turner, J. S. *J. geophys. Res.* **85**, 2573–2593 (1980).
6. Turner, J. S. *Nature* **285**, 213–215 (1980).
7. Irvine, T. N. in *Physics of Magmatic Processes* (ed. Hargraves, R. B.) 325–383 (Princeton University Press, 1980).
8. Huppert, H. E. & Sparks, R. S. J. *Nature* **286**, 46–58 (1980); *Contr. Miner. Petrol.* **75**, 279–289 (1980).
9. Huppert, H. E. & Turner, J. S. *Earth planet. Sci. Lett.* **54**, 144–152 (1981).
10. Usselman, T. M. & Hodge, D. S. *J. Volcanol. geotherm. Res.* **4**, 265–281 (1978).
11. Busse, F. H. *Rep. Prog. Phys.* **41**, 1930–1967 (1978).
12. Kraichnan, R. H. *Phys. Fluids* **5**, 1374–1389 (1962).
13. Jaupart, C. thesis, Massachusetts Instit. Technol. (1981).
14. White, D. thesis, Univ. Cambridge (1982).
15. Richter, F. M., Nataf, H. C. & Daly, S. F. *J. Fluid Mech.* **129**, 173–192 (1983).

16. Murase, T. & McBirney, A. R. *Geol. Soc. Am. Bull.* **84**, 3563–3592 (1973).
17. Richet, P. & Bottinga, Y. *Geochim. cosmochim. Acta* **44**, 1535–1541 (1980).
18. Bottinga, Y., Richet, P. & Weill, D. F. *Bull. Miner.* **106**, 129–138 (1983).
19. Urbain, G., Bottinga, Y. & Richet, P. *Geochim. cosmochim. Acta* **46**, 1061–1072 (1982).
20. Merzkirch, W. *Flow Visualization* (Academic, New York, 1974).
21. Sparrow, E. M., Husar, R. B. & Goldstein, R. J. *J. Fluid. Mech.* **41**, 793–800 (1970).
22. Turner, J. S. *Buoyancy Effects in Fluids* (Cambridge University Press, 1973).
23. Gill, A. E. J. *Fluid Mech.* **26**, 515–536 (1966).
24. Baines, W. D. & Turner, J. S. *J. Fluid Mech.* **37**, 51–80 (1969).
25. Donaldson, C. H. *Miner. Mag.* **41**, 323–336 (1977).
26. Samoylovitch, Y. A. *Geochem. Int.* **16**, 79–84 (1979).
27. Brandeis, G., Jaupart, C. & Allègre, C. J. *J. geophys. Res.* (submitted).

Isotope composition of sulphate in acid mine drainage as measure of bacterial oxidation

Bruce E. Taylor & Mark C. Wheeler*

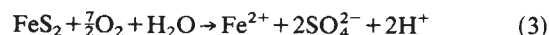
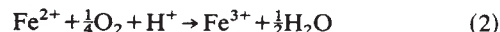
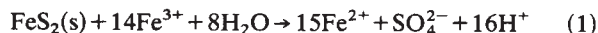
Department of Geology, University of California, Davis, California 95616, USA

Darrel Kirk Nordstrom

Water Resources Division, US Geological Survey, Menlo Park, California 94025, USA

The formation of acid waters by oxidation of pyrite-bearing ore deposits, mine tailing piles, and coal measures is a complex biogeochemical process and is a serious environmental problem. We have studied the oxygen and sulphur isotope geochemistry of sulphides, sulphur, sulphate and water in the field and in experiments to identify sources of oxygen and reaction mechanisms of sulphate formation. Here we report that the oxygen isotope composition of sulphate in acid mine drainage shows a large variation due to differing proportions of atmospheric- and water-derived oxygen from both chemical and bacterially-mediated oxidation. ¹⁸O-enrichment of sulphate results from pyrite oxidation facilitated by *Thiobacillus ferrooxidans* in aerated environments. Oxygen isotope analysis may therefore be useful in monitoring the effectiveness of abatement programmes designed to inhibit bacterial oxidation. Sulphur isotopes show no significant fractionation between pyrite and sulphate, indicating the quantitative insignificance of intermediate oxidation states of sulphur under acid conditions.

The overall stoichiometry of pyrite oxidation may be described by the reaction: $\text{FeS}_2(\text{s}) + \frac{15}{4}\text{O}_2(\text{aq}) + \frac{7}{2}\text{H}_2\text{O} \rightarrow \text{Fe}(\text{OH})_3(\text{s}) + 2\text{H}_2\text{SO}_4(\text{aq})$. In order to understand the oxidation kinetics, intermediate steps in this reaction must be considered¹. The following three reactions have been shown to operate under acid conditions^{2–5}:

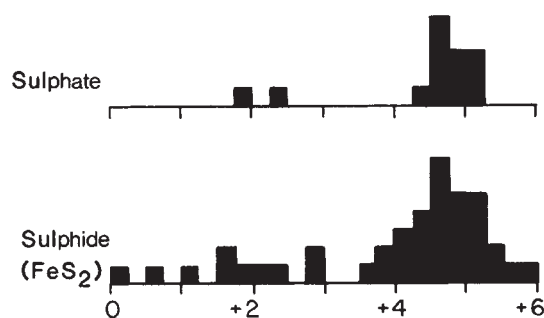


Below a pH of 3, reaction (2) is thought to be the rate-limiting step for reaction (1), the principal inorganic oxidation mechanism³. *Thiobacillus ferrooxidans* increases the rate of oxidation of ferrous to ferric iron (reaction (2)) by several orders of magnitude in acid environments^{4,6,7}. This microbial influence on pyrite oxidation is usually referred to as indirect action, as distinguished from a direct action mechanism in which *T. ferrooxidans* or other microbes attach themselves to the pyrite surface^{8,9} and directly attack the surface, enzymatically oxidizing sulphide to sulphate, similar to reaction (3)⁵.

The roles and sources of oxygen may thus be quite varied. In reaction (1) the sulphate oxygen is derived from the water molecule whereas in reaction (3) it is derived from dissolved

* Present address: US Geological Survey, 345 Middlefield Road, Menlo Park, California 94025, USA.

WEST SHASTA DISTRICT



LEVIATHAN MINE

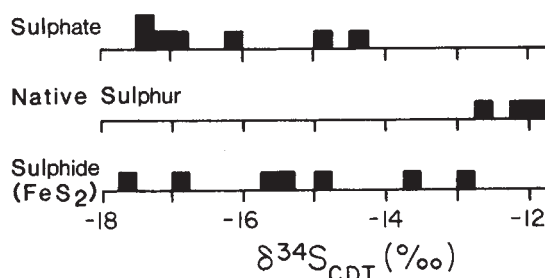


Fig. 1 Comparison of sulphur isotope compositions of primary sulphide and sulphate produced during sulphide oxidation. The similarity of $\delta^{34}\text{S}$ values indicates little or no sulphur isotope fractionation during pyrite oxidation, suggesting that sulphate is the dominant aqueous sulphur species. Intermediate oxidation states of sulphur, although important in the reaction mechanism, are quantitatively small. Each square represents one analysis.

oxygen (87.5%) and water (12.5%). Each of the proposed reactions may contribute to the dissolution of pyrite, according to different rates and to the influence of many poorly understood microenvironmental factors¹. The marked difference in the oxygen isotope composition of atmospheric oxygen ($\delta^{18}\text{O} = +23.0\text{‰}$)¹⁰ and meteoric waters ($\delta^{18}\text{O} = -9.5$ to -15.5‰ in our study) offers the possibility of tracing principal oxidation mechanisms by isotopic analysis of the sulphate formed. Because the rate of oxygen isotope exchange between sulphate and water is slow, even at a low pH^{11,12}, the oxygen isotope composition of sulphate formed during pyrite oxidation should preserve the identity of the original source of oxygen.

We have determined the oxygen and sulphur isotope compositions of dissolved sulphate and water in acid mine drainages in the West Shasta Cu-Zn district and at the Leviathan mine, California; Argo tunnel, Colorado; and Alpine Gulch area, San Juan Mountains, Colorado. The most abundant sulphide mineral in these deposits is pyrite (FeS_2). Sulphur was available for analysis only at the Leviathan mine.

Water samples were filtered in the field using a 0.1 μm filter and portable pump. After removal of cations on a BioRad Ag 50W-X12 resin bed in the laboratory, sulphate was precipitated as BaSO_4 in a heated BaCl_2 solution. The oxygen isotope composition of the sulphate was determined by mass spectrometry on CO_2 prepared by combustion of BaSO_4 with pure graphite enclosed in molybdenum foil and heated *in vacuo* by induction in a molybdenum crucible. Quantitative yields were attained by conversion of any CO to CO_2 by discharge between platinum electrodes¹³ exposed to the BaSO_4 combustion vessel. Water was analysed for oxygen isotopes by equilibration with CO_2 at 25 °C using a fractionation factor of 1.0412 (ref. 14). SO_2 for sulphur isotope analysis was prepared from sulphides, and from BaSO_4 ground and mixed together with quartz glass¹⁵, by combustion with CuO.

Isotopic data are reported in the δ -notation, where δ (in per mille, ‰) $= [(R_{\text{SA}} - R_{\text{STD}})/R_{\text{STD}}] \times 10^3$ and $R = ^{18}\text{O}/^{16}\text{O}$ or

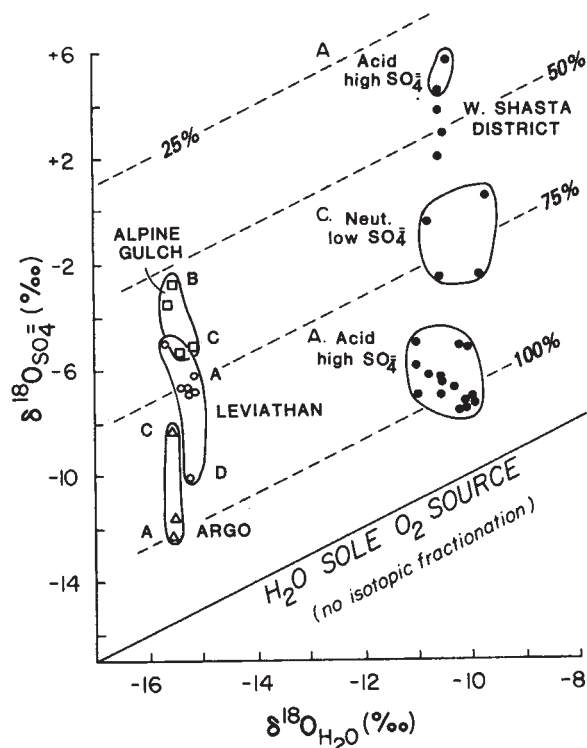


Fig. 2 Oxygen isotope compositions of acid water sulphate relative to the isotopic compositions of local meteoric water for four study areas. Samples represent analyses of mine portal effluents (the most ^{18}O -depleted of the Argo and West Shasta samples), polluted streams and seeps, and non-polluted streams; A, acid, high SO_4^{2-} ($\geq 500 \text{ mg l}^{-1}$); B, acid, moderate SO_4^{2-} (50–500 mg l^{-1}); C, neutral, low SO_4^{2-} ($\leq 50 \text{ mg l}^{-1}$); D, neutral, moderate SO_4^{2-} . pH values as low as 1.6 and sulphate concentrations as high as 800 mg l^{-1} were measured in mine effluents in the West Shasta district. The 'H₂O sole O₂ source' line describes zero oxygen isotope fractionation between the sulphate and water. The broken lines for 100, 75, 50 and 25% indicate the contributions of reaction 1 to pyrite oxidation and the expected oxygen isotope compositions of the product sulphate based on fractionations determined in experiments described in the text.

$^{34}\text{S}/^{32}\text{S}$ (SA refers to sample, STD to standard). $\delta^{18}\text{O}$ (relative to V-SMOW) were reproducible to 0.1‰ for sulphate, and 0.2‰ for water. $\delta^{34}\text{S}$ values (relative to CDT) were reproducible to 0.05‰.

The sulphur isotope composition of sulphate in acid mine drainage, fresh pyrite, and sulphur from the W. Shasta district and the Leviathan mine are shown in Fig. 1. The isotopic data are compatible with formation of Leviathan mine sulphur and sulphide from the same ^{34}S -depleted hydrothermal H_2S , but under non-equilibrium, variably oxidizing conditions. Laboratory oxidation of aqueous sulphide by *Thiobacillus* has yielded depletions in ^{34}S in the oxidized products: 6% in sulphur^{16,17}, and up to 18% in sulphate¹⁷. When pyrite was used as a substrate, a mixed culture of microbes produced sulphate depleted in ^{34}S by only 1.7% relative to the original pyrite¹⁸ and no difference was detected in experiments using *T. ferrooxidans*. Field¹⁹ found little or no sulphur isotope fractionation between primary sulphide and secondary minerals during the oxidation of pyrite-bearing ore deposits, and Schoen and Rye²⁰ also observed little or no fractionation of sulphur isotopes upon oxidation of H_2S to form native sulphur and sulphate minerals in solfataras at Yellowstone National Park. From Fig. 1, we conclude that the sulphur isotopes are not measurably fractionated during pyrite oxidation. The similarity of sulphur isotope compositions of sulphide and dissolved sulphate shown in Fig. 1, particularly for West Shasta, corroborates earlier studies of sulphide mineral oxidation. Further, the lack of isotopic fractionation indicates that, while required in the oxidation pathway²¹, aqueous sulphur species of intermediate oxidation states are quantitatively small,

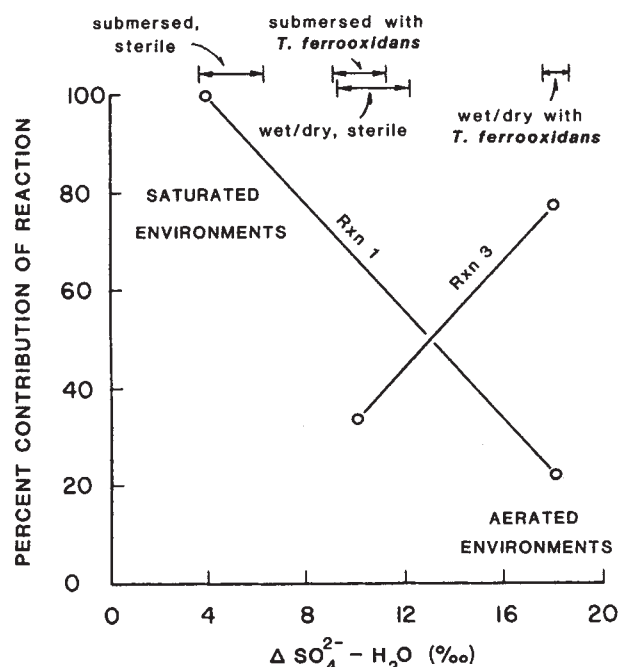


Fig. 3 Per cent contributions of reaction for oxidation reactions (1) and (3) plotted against expected oxygen isotope fractionation between sulphate and water. End-member points (open circles) are plotted for extreme fractionations measured in the following conditions: anaerobic, sterile and aerobic with *T. ferrooxidans* present. Also shown are the ranges for this fractionation (^{18}O enrichment, $\Delta\text{SO}_4^{2-} - \text{H}_2\text{O}$) observed in experiments using submersed and alternating wet/dry conditions with and without *T. ferrooxidans*. Saturated environments would be represented by flooded mines, whereas aerated environments (characterized by an increase in utilization of molecular oxygen via reaction (3)) would be found in stream beds and the near-surface unsaturated zone.

although they are certainly important²¹. Thus, the sulphur isotope composition of sulphate can indicate the source of the sulphur in acid waters. For example, although the range of $\delta^{34}\text{S}$ values for sulphide at the Leviathan mine may partly reflect microscopic contamination with sulphur, the acid effluents are indicated to form primarily from oxidation of pyrite rather than sulphur.

Figure 2 illustrates a large variation for the oxygen isotope composition of sulphate in acid mine drainage from several sites. A 14% variation is shown for samples of acid mine drainage from portal effluents and heavily polluted streams in the West Shasta district. The oxygen isotope enrichment of sulphate relative to water approaches a minimum of 3‰ for acid mine effluents from the Argo tunnel (Colorado) and several mines in the West Shasta district. Isotopically heavier sulphate was found in waters draining more aerated hydrological environments, such as high sulphate, acid streams in the West Shasta district ($\delta^{18}\text{O} = +2$ to $+6\%$; Fig. 2), or the exposed bog (seep) environments at Alpine gulch. The neutral waters from West Shasta represent unpolluted streams with low concentrations ($\leq 50 \text{ mg l}^{-1}$) of sulphate probably derived from sources other than the oxidizing pyrite ore bodies. Oxygen isotope enrichments of sulphate with respect to meteoric water in these non-polluted streams are approximately $+9.5\%$ at West Shasta and $+7.5\%$ near the Argo tunnel site. Since the $\delta^{18}\text{O}_{\text{SO}_4^{2-}}$ values of sulphate in neutral waters shown in Fig. 2 are within the range of $\delta^{18}\text{O}$ values for acid mine water sulphate, processes other than simple mixing of streams and mine effluents are involved. Different sources of oxygen, rates and mechanisms of oxidation, variation in isotopic fractionations, and sites of activity of *T. ferrooxidans* or other sulphur-oxidizing microbes must be considered.

To determine the isotopic composition of dissolved oxygen utilized during the oxidation of pyrite, we conducted experiments similar in design to those of Lloyd¹². Lloyd determined the isotopic compositions of aliquots of a pure oxygen atmo-

sphere periodically withdrawn from closed vessels in which aqueous Na_2S was undergoing oxidation. The details of our experiments will be reported elsewhere. We found, as did Lloyd, that molecular oxygen used in sulphide oxidation is depleted in ^{18}O relative to the initial atmospheric oxygen. In our experiments, sulphate rather than sulphite was the oxidized end product, and was isotopically heavier than predicted from Lloyd's results. The kinetic isotope fractionation factor may be determined by plotting the fraction of oxygen remaining (log scale) versus the change in isotopic composition of the oxygen (linear scale). This plot yields a straight line, the slope of which is the kinetic fractionation factor¹². Of course, this factor must reflect the combined isotopic enrichments or depletions attending both reactions (2) and (3). From our experiments, the values for the kinetic fractionation of atmospheric oxygen during pyrite oxidation are -4.6% for inorganic oxidation pathways, and -11.12% for oxidation facilitated by *T. ferrooxidans*. Thus, in natural environments, pyrite oxidation could use molecular oxygen with $\delta^{18}\text{O} = +18.4\%$ via inorganic mechanisms (reactions (2) and (3)), and $\delta^{18}\text{O} = +11.8\%$ in *T. ferrooxidans*-mediated processes (principally, reaction (3), by direct attack).

A different set of experiments was designed to mimic the submersed conditions of flooded mines, and also the alternating wet/dry conditions of tailings piles or dumps. These experiments were open to the atmosphere and were run either sterile or inoculated with *T. ferrooxidans*. Submersed experiments yielded oxygen isotope enrichments of sulphate of $+4.1$ to $+6.2\%$ relative to water under sterile conditions, and $+8.9$ to $+10.9\%$ when *T. ferrooxidans* was present. Alternating wet/dry experiments produced isotopically heavier sulphate, with enrichments of $+9.5$ to $+12.1\%$ (sterile) and $+17.6$ to $+18.1\%$ (with *T. ferrooxidans*). Both exposure of the pyrite to atmospheric oxygen during partial drying and the presence of *T. ferrooxidans* resulted in a marked ^{18}O enrichment of the sulphate. This is most reasonably explained by facilitation of reaction (3). In the sterile experiments, this would happen by reducing the contact time of the water with pyrite, favouring reaction (3). *T. ferrooxidans* may directly attack pyrite, oxidizing it via a pathway analogous to reaction (3), thereby incorporating molecular oxygen in the sulphate.

The natural variations of $\delta^{18}\text{O}_{\text{SO}_4^{2-}}$ in Fig. 2 must reflect the incorporation of both water-derived and atmospheric oxygen. This could occur in different environments which would favour either reaction (1) or reaction (3). For example, where dissolved oxygen is increased in aerated environments, or where *T. ferrooxidans* or similar microbes are present, reaction (3) may be favoured (that is, mediated), increasing the $\delta^{18}\text{O}$ of sulphate formed by incorporation of isotopically heavy atmospheric oxygen. Schwarcz and Cortecchi²² suggested that the sulphate should contain approximately equal amounts of water and atmospheric oxygen. Considering reactions (1) and (3) as the principal oxidation pathways, the results of our experiments suggest that the variations in $\delta^{18}\text{O}_{\text{SO}_4^{2-}}$ in Fig. 2 may be modelled in terms of a percentage contribution of these reactions to the total sulphate. The percentage of sulphate formed via reaction (1), where all oxygen is derived from the water, is plotted in Fig. 2. The various contributions of reactions (1) and (3) are plotted against the isotopic enrichment of sulphate relative to water in Fig. 3, based on the results of experiments described above. Enrichments of 3–5‰ in effluents from flooded mines (submersed environment) suggest inorganic oxidation of pyrite via reaction (1), with less than 10% of the sulphate oxygen derived from the atmosphere. Isotopic enrichments of sulphate up to 16‰ are compatible with *T. ferrooxidans*-mediated oxidation of pyrite as the primary source of sulphate. Here, the contribution of molecular oxygen to sulphate approaches 70%. Thus, the isotopic analysis of sulphate and water may implicate the involvement of *T. ferrooxidans* in the mediation of pyrite oxidation, and may be useful in monitoring the effectiveness of abatement programmes designed to inhibit bacterial oxidation.

This research was supported by NSF grant EAR-79-11144. We thank Mr Dennis Murphy for samples from Alpine Gulch

and Dr Aaron Mills for *T. ferrooxidans*. Mr Alan Welch and Mr Dale Hammermeister (USGS) kindly provided funding for samples obtained from the Leviathan mine area. Illustrations were drafted by E. Bailey.

Received 19 September 1983; accepted 12 January 1984.

1. Nordstrom, D. K. in *Acid Sulfate Weathering*, 37–56 (Soil Science Society of America, Madison, 1982).
2. Garrels, R. M. & Thompson, M. E. *Am. J. Sci.* **258**, 57–67 (1960).
3. Singer, P. C. & Stumm, W., *Second Symp. Coal Mine Drainage Res.*, 12–34 (Mellon Inst., Pittsburgh, 1968).
4. Singer, P. C. & Stumm, W. *Science* **167**, 1121–1123 (1970).
5. Ehrlich, H. L. *Geomicrobiology*, (Dekker, New York, 1981).
6. Lacy, D. T. & Lawson, F. *Biotech. Bioengng.* **12**, 29–50 (1970).
7. Nordstrom, D. K. thesis, Stanford Univ., Calif. (1977).
8. Tuovinen, O. H. & Kelly, D. P. *Z. allg. Mikrobiol.* **12**, 311–346 (1972).
9. Bennett, J. C. & Tributsch, H. *J. Bact.* **134**, 310–317 (1978).
10. Kroopnick, P. & Craig, H. *Science* **175**, 54–55 (1972).
11. Hoering, T. C. & Kennedy, J. W. *J. Am. chem. Soc.* **79**, 56–60 (1967).
12. Lloyd, R. M. *J. geophys. Res.* **73**, 6099–6110 (1968).
13. Nehring, N. L., Bowen, P. A. & Truesdell, A. H. *Geothermics* **5**, 63–66 (1977).
14. Friedman, I. & O'Neil, J. R. *U.S. Geol. Survey Prof. Pap.* 440-KK (1977).
15. Coleman, M. L. & Moore, M. P. *Analyt. Chem.* **50**, 1594–1595 (1978).
16. Kaplan, I. R. & Rittenburg, S. C., *J. gen. Microbiol.* **34**, 195–212 (1964).
17. Kaplan, I. R. & Rafter, T. A. *Science* **127**, 517–518 (1958).
18. Nakai, N. & Jensen, M. L. *Geochim. cosmochim. Acta* **28**, 1893–1912 (1964).
19. Field, C. W. *Econ. Geol.* **61**, 1428–1435 (1966).
20. Schoen, R. & Rye, R. O. *Science* **170**, 1082–1084 (1970).
21. Goldhaber, M. B. *Am. J. Sci.* **283**, 193–217 (1983).
22. Schwarcz, H. P. & Cortecci, G., *Chem. Geol.* **13**, 285k–294k (1974).

Grooming, alliances and reciprocal altruism in vervet monkeys

Robert M. Seyfarth & Dorothy L. Cheney

Department of Anthropology, University of California, Los Angeles, California 90024, USA, and Institute of Primate Research, National Museums of Kenya, PO Box 34505, Nairobi, Kenya

Reciprocal altruism refers to the exchange of beneficial acts between individuals, in which the benefits to the recipient exceed the cost to the altruist. Theory predicts that cooperation among unrelated animals can occur whenever individuals encounter each other regularly and are capable of adjusting their cooperative behaviour according to experience¹. Although the potential for reciprocal altruism exists in many animal societies, most interactions occur between closely related individuals, and examples of reciprocity among non-kin are rare^{2,3}. The field experiments on vervet monkeys which we present here demonstrate that grooming between unrelated individuals increases the probability that they will subsequently attend to each others' solicitations for aid. Vervets appear to be more willing to aid unrelated individuals if those individuals have behaved affiliatively toward them in the recent past. In contrast, recent grooming between close genetic relatives appears to have no effect on their willingness to respond to each other's solicitations for aid.

Grooming is the most common form of affiliative behaviour in non-human primates, and occurs when one animal picks through the fur of another, removing any ectoparasites found. The formation of alliances, in which two individuals jointly direct aggression against a third, is also common. In such Old World monkey species as baboons, macaques and vervets, where these behaviours have been studied extensively, most grooming and alliances involve animals that are closely related through the material line (paternity is usually unknown). Their evolution has, therefore, partly been explained in terms of kin selection^{4–7}. Similar affiliative interactions, however, also occur between non-relatives. Such cooperative behaviour could establish long-term social bonds from which the individual could derive future reciprocal benefit^{8,9}. Empirical evidence of reciprocal altruism, however, has thus far been limited to two studies of non-human primates, which have suggested that unrelated animals often form the most alliances with those individuals who also support them^{2,3}. Other studies have suggested that grooming among

non-kin may also increase the probability of future benefit, either in the form of alliances, tolerance at feeding sites, or reduced aggression^{10–14}. Although there are data to support this hypothesis, no study has yet demonstrated a causal relationship between grooming and future cooperative behaviour among non-kin.

To test the hypothesis that grooming may increase the probability of future support in aggressive disputes, we designed the present experiment, using as our subjects adult female and juvenile members of three groups of free-ranging vervet monkeys (*Cercopithecus aethiops*) in Amboseli National Park, Kenya. First, two individuals, A and B, were selected for a matched pair of trials. We then made tape-recordings of a vocalization used by animal A to solicit support from others in an alliance. This vocalization was then played near animal B in one of two conditions: either after A had groomed B or after a period of time when no grooming had occurred. Our aim was to determine whether prior grooming increased the likelihood that B would attend to A's solicitation, and hence to establish whether grooming and alliances formed a reciprocal system of interactions in vervet monkeys.

The vocalization used in playbacks was one given by vervets when threatening another individual and simultaneously soliciting support from those nearby¹⁵. Previous observations showed that this vocalization occurs in conjunction with visual threat displays and is limited to circumstances when the vocalizer and its opponent are in visual contact. It is, therefore, unlikely that animal B would interpret the vocalization as directed against itself. Instead, playback of animal A's threat vocalization mimicked a situation in which A solicited B's support in an alliance.

A's vocalization was played to B at least 30, but not more than 90 min, after A had groomed B for at least 1 min. In the control A's vocalization was played to B after a period when A had not groomed B during the previous 2 h or more. The vocalization was played from a concealed speaker placed at a mean distance of 7.7 m from B (s.d. = 2.6). To ensure that playbacks mimicked vocalizations given in natural conditions, we conducted trials in bushy areas where vervets frequently vocalize when out of sight of one another. The speaker was always placed in an area where the vocalizer was likely to have been at the time. All trials, however, were conducted when the vocalizer was out of sight of the subject. To ensure further that the subject would not interpret playback as aggression directed against itself, we varied the speaker's position so that it did not always face animal B. Animal B was filmed for 10 s before playback to establish the probability that it would look towards

Table 1 Effect of grooming by vocalizer on duration of response shown by subjects to playback of recruitment vocalization

Vocalizer	Subject	Duration of response (s)	
		Prior grooming	No prior grooming
J PC	J AF	3.7	1.3
J AF	J PC	1.0	1.1
A TY	J SN	7.3	0.6
J SN	A TY	6.8	1.3
J HB	J CC	7.2	3.4
J PC	A AM	9.3	2.5
A TY	A NT	0.9	0
J AO	A CY	8.1	0.7
J AO	J CO	6.7	0.4

Response was looking towards the speaker. Duration of looking is defined in text, and is measured in frames of film at 18 f.p.s. All individuals except CO are females; A = adult, J = juvenile (aged <4 yr). Two-tailed Wilcoxon test, $n = 9$, $T = 1$, $P < 0.01$. Results exclude one pair of trials involving adult females in which the film in one trial was too dark to permit accurate analysis. Spoken commentary made during the trial, however, indicated that duration of response after grooming was 2 s longer than after no prior grooming. If this pair of trials is included, $n = 10$, $T = 1$, $P < 0.01$.

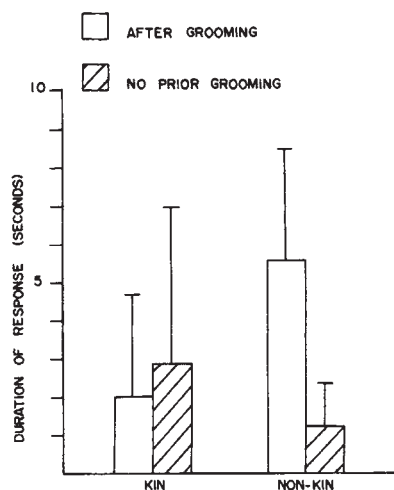


Fig. 1 Means and standard deviations for responses of subjects to the recruitment vocalizations of kin and non-kin in the presence and absence of prior grooming by the vocalizers. Grooming by non-kin elevated the responses of subjects over those of kin as follows: non-kin after grooming versus kin after grooming, two-tailed Mann-Whitney U -test, $n_1 = 10$, $n_2 = 9$, $U = 12.5$, $P < 0.05$; non-kin after grooming versus kin after no grooming; $U = 21$, $P < 0.10$. Since cross-group comparison did not control for individuals, the Mann-Whitney U -test was used rather than the Wilcoxon matched-pairs signed ranks test. No other comparison across kin and non-kin groups were significant. See text for within-group comparisons.

the speaker (or show any other response) in the absence of any call. Filming continued for an additional 10 s after the call had ended. Animal B's willingness to attend to A's solicitation was measured by the intensity of B's response to A's vocalization. If grooming increased the probability of future support in aggressive disputes, we predicted that B's response would be stronger when it had recently been groomed by A than when no grooming had recently occurred. We assumed that vervets recognize each others' vocalizations individually, as an earlier study¹⁶ showed that mothers look towards screams of their offspring for significantly longer times than did unrelated females.

Long-term data on genetic relatedness through the maternal line were available for most of the vervet monkeys in the three social groups. We were therefore able to test for reciprocity both among pairs of kin ($r \geq 1/4$ through the maternal line) and among pairs of non-kin ($r < 1/8$). Nine kin pairs and 10 non-kin pairs served as subjects. Among non-kin, we controlled for dominance rank by taking an equal number of subjects who were dominant or subordinate to the animal whose call was being played. In both groups, an equal number of subjects was tested first after grooming and first in the control condition, without prior grooming. Five individuals appeared as subjects in both the kin and non-kin groups. Within each group, however, individuals appeared as subjects only once. Successive trials involving the same subject were always separated by over 48 h. Films were analysed using a frame-by-frame editor, and were supplemented by commentary spoken into a tape recorder at the time of the experiment. To control for each subject's behaviour at the time a trial was conducted, responses were defined in terms of the duration of a given behaviour in the 10 s after playback minus the duration of the same behaviour in the 10 s before playback¹⁷.

As in earlier experiments¹⁶, subjects responded to playbacks by looking toward the speaker. When no prior grooming had occurred, the responses of kin were generally stronger than those of non-kin, although not significantly so (Fig. 1). Prior grooming had no significant effect on the behaviour of kin, who responded to each others' solicitations for aid with similar intensity, regardless of recent affiliative interactions (two-tailed Wilcoxon test, $n = 9$, one tie, $T = 12$, $P > 0.10$). Prior grooming

had a strong effect, however, on non-kin. In trials involving non-kin, the duration of subjects' responses was significantly longer after grooming had occurred than when it had not (Table 1). The effect of prior grooming on the behaviour of non-kin was so marked that it often increased the responses of non-kin to a level above those shown by kin (Fig. 1). Thus among non-kin, grooming of animal B by animal A significantly increased the intensity of B's response to A's solicitations for support. In contrast, recent grooming did not increase the willingness of kin to attend to each other's disputes.

For cooperative behaviour to evolve through reciprocal altruism individuals must modify their behaviour depending on who has behaved altruistically toward them in the past^{1,18}. Our present results support the theory of reciprocal altruism, and provide experimental evidence that one (but by no means the only) function of grooming among non-kin is to increase the responsiveness of others to the groomer's subsequent solicitations for aid. The fact that playbacks elicited only looking responses suggests, moreover, that, although grooming may increase the readiness of individuals to attend to each others' solicitations for support, actual physical involvement may depend on further assessment of the potential costs of intervention, such as the relative dominance ranks of the participants and the intensity of the dispute^{16,17,19}.

There are several possible explanations for the behaviour of kin in these experiments. High rates of affiliative interactions among kin may reduce the extent to which an individual's behaviour depends on interactions that have occurred in the immediate past. Thus the occurrence or lack of a single grooming bout may have little effect on relationships among kin, as opposed to those among non-kin. Moreover, bonds formed among kin during infancy may have long-term consequences that override the effects of many subsequent social interactions. Whatever the proximate mechanisms involved, kin selection theory²⁰ predicts that an individual can increase its fitness by aiding its relatives regardless of whether prior altruistic behaviour has occurred. This prediction is entirely consistent with our results presented here.

The experiments described here provide support for the theory of reciprocal altruism by showing that a vervet monkey increases the likelihood that an unrelated individual will attend to its solicitations for aid if it has behaved affiliatively towards that individual in the recent past. In contrast, support from close genetic relatives appears to be relatively independent of recent affiliative interactions.

We thank the Office of the President and the Ministry of Tourism and Wildlife of the Republic of Kenya for permission to conduct research in Amboseli National Park. We also thank B. Oguya, the Warden of Amboseli, and J. Else, Director of the Institute of Primate Research in Nairobi, for their help in all aspects of field research. We also thank B. Smuts and R. Wrangham for their comments on the manuscript. Research was supported by the H. F. Guggenheim Foundation and NSF grant BNS 82-15039. The manuscript was written while the authors held fellowships at the Center for Advanced Study in the Behavioural Sciences, where we acknowledge support from the NSF (BNS 76-22943), the Alfred P. Sloan Foundation (82-2-10) and the Exxon Education Foundation.

Received 11 October 1983; accepted 17 January 1984.

1. Trivers, R. L. *Q. Rev. Biol.* **46**, 35-57 (1971).
2. Packer, C. *Nature* **265**, 441-443 (1977).
3. de Waal, F. *Chimpanzee Politics* (Harper & Row, New York, 1982).
4. Cheney, D. L. *Behav. Ecol. Sociobiol.* **2**, 303-318 (1977).
5. Cheney, D. L. in *Primate Social Relationships* (ed. Hinde, R. A.) (Blackwell, Oxford, 1983).
6. Kurland, J. A. *Contr. Primatol.* **12**, 1-145 (1977).
7. Massey, A. *Behav. Ecol. Sociobiol.* **2**, 31-40 (1977).
8. Seyfarth, R. M. *J. theor. Biol.* **65**, 671-698 (1977).
9. Chalmers, N. *Social Behaviour in Primates* (Arnold, London, 1979).
10. Stambach, E. *Behaviour* **67**, 322-338 (1978).
11. Chapais, B. & Schulman, S. R. *J. theor. Biol.* **82**, 47-89 (1980).
12. Fairbanks, L. *Anim. Behav.* **28**, 853-859 (1980).
13. Seyfarth, R. M. *Anim. Behav.* **28**, 798-813 (1980).
14. Silk, J. B. et al. *Behaviour* **78**, 111-137 (1981).
15. Struhsaker, T. T. in *Social Communication among Primates* (ed. Altmann, S.) 281-324 (University of Chicago Press, 1967).

16. Cheney, D. L. & Seyfarth, R. M. *Anim. Behav.* **28**, 362–367 (1980).
 17. Cheney, D. L. & Seyfarth, R. M. *Anim. Behav.* **30**, 739–751 (1982).
 18. Axelrod, R. & Hamilton, W. D. *Science* **211**, 1390–1396 (1981).
 19. Gouzoules, S. *et al. Anim. Behav.* **32**, 182–193 (1984).
 20. Hamilton, W. D. *J. theor. Biol.* **7**, 1–52 (1964).

Earliest known proboscidean from early Eocene of north-west Africa

M. Mahboubi*, R. Ameer*, J. Y. Crochet† & J. J. Jaeger‡

* ONRS-CERD, Département de Géologie, Université d'Oran, BP16, Oran-Es Sénia, Algeria

† LA 327, CNRS Institut des Sciences de l'Evolution, Université des Sciences et Techniques du Languedoc (Montpellier II), place Eugène Bataillon, 34060 Montpellier-Cédex, France

‡ ERA 963, CNRS Laboratoire de Paléontologie des Vertébrés et de Paléontologie Humaine, tour 25–15, Université Pierre et Marie Curie (Paris VI), 4 place Jussieu, 75230 Paris-Cédex, France

The earliest known proboscidean remains have now been found at a new early Eocene locality at Brezina in southern Algeria (El Kohol). These new finds, represented by complete skulls and postcranial material, show several unexpected derived characters shared with the modern representatives of the Elephantoidea and the Deinotheriidae, suggesting close phylogenetic affinities and demonstrating also the great antiquity of the differentiation of modern proboscideans in Africa. These remains have been dated by associated charophyte flora and vertebrate remains which constitute the oldest known vertebrate community from the African Eocene.

The proboscidean remains are of an animal not exceeding a height of about 1 m, with a proportionally large head, and with graviportal limbs as in modern elephants. The skull is characterized by its considerable height, the presence of a conspicuous transversal frontomaxillary bulge, and an important pneumatization (Figs 1a, b, d and 2). The nasal opening is located at the anterior end of the skull. The orbit is in an anterior and low position; the occipital region is elevated. The orbit is largely confluent with the temporal fossa and bordered by a maxilla which extends upward and forward. It is limited posterodorsally by the postorbital apophysis of the frontal and ventrally by a zygomatic apophysis of the maxilla, the ventral face of which is excavated by a deep groove-like fossa. The latter extends from the posterior part of the premaxilla to the level of M2. The lachrymal is not observable on the skull shown, but it has been observed on another specimen, although as a rule the sutures are obliterated. The temporal ridges start from the postorbital apophyses of the frontals, do not reach each other, and are continued rearward by powerful nuchal crests. The base of the zygomatic arch bears a strong downward directed process. The

arch then extends obliquely upward to reach the external edge of the glenoid cavity (Fig. 2). The external auditory meatus, partially roofed by an apophysis of the squamosal, opens behind the glenoid cavity.

The upper dentition is almost complete, P1 alone has disappeared; I2 is the most developed of the three incisors, while the others are reduced (Fig. 1b). The canine shows a development comparable to that of I3. The upper and lower cheek teeth are very similar to those of *Barytherium grave* (Andrews 1901)^{1,2} but of definitely smaller size.

On the lower jaw (Fig. 1c) there are only two incisors and no canine. The central incisors (I2?) are strongly procumbent and spatulate and I3 is reduced. The horizontal ramus of the mandible shows a long symphysis extending to the level of the P4. The ascending ramus is subvertical. Its anterior edge hides the talonid of M3.

The structure of the girdles shows that it is a fully terrestrial animal. The limb bones have many derived characters common to both the Elephantoidea and the Deinotheriidae, associated with primitive characters such as the presence of an entepicondylar foramen on the humerus, a character hitherto unreported in a proboscidean. The forelimbs are extremely robust. The humerus and the bones of the forearm are of similar lengths. However, the femur is longer than the humerus and much longer than the tibia. The knee is thus in a very distal position. The astragalus and the calcaneum are nearly identical with those of *Paleomastodon beadnelli* (Andrews 1901), the earliest known Elephantoidea³. These resemblances provide crucial evidence for considering this form as close to both the Elephantoidea and the Deinotheriidae.

Many derived characters indicate that this form belongs to the Proboscidea⁴ and include: (1) the anterior position of the orbit; (2) the strong development of I2; (3) the high position of the external auditory duct and the glenoid cavity; (4) the partial closure of the external auditory duct by a post-tympanic apophysis of the squamosal; (5) the absence of a condylar foramen; (6) the general shape and structure of the astragalus.

The order Proboscidea also includes more primitive groups characteristic of the late Eocene and Oligocene of Africa, represented by the genera *Moeritherium* and *Barytherium* (Andrews 1901)^{1–3,5}. *Moeritherium* shares some primitive features with the new Algerian form: (1) a comparable dental formula; (2) the morphology of the cheek teeth; (3) the slope of the palate and of the premaxilla.

However, the Algerian form differs from *Moeritherium* by such derived characters as the elevation and the pneumatization of the skull and the presence of a transversal frontomaxillary bulge, while *Moeritherium* exhibits a number of other derived features suggesting an adaptation to a semi-aquatic life.

The Algerian form resembles *Barytherium grave* (Andrews 1901) in molar structure and the weak molarization of its premolars. However, the many observable differences between these two forms seem to preclude a direct ancestor–descendant relationship. In particular, the deep submaxillary fossa described

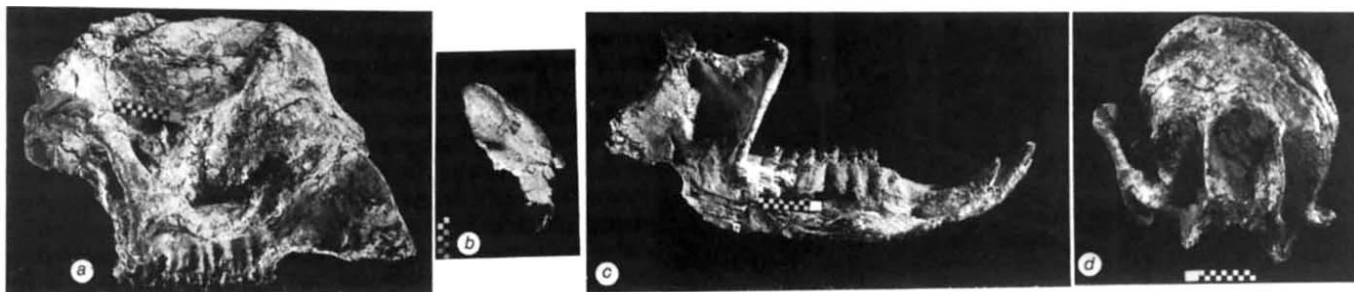


Fig. 1 a, Skull (KAI-18) of the proboscidean from El Kohol, in lateral view; there is no anterior tooth preserved on that skull. b, Fragment of right premaxilla (KA X-16) showing upper I2 strongly developed. c, Right lower jaw (KA X-11) of the proboscidean from El Kohol in lateral view, showing the strong development of the central incisor (I2?), the diastema and the cheek teeth. P2 is missing. d, Skull of the proboscidean from El Kohol in anterior view (KAI-18) (photos C. Abrial).

in the Algerian form does not seem to be present in *B. grave*. Moreover, the structure of the astragalus of the latter species is distinct from that of other known Proboscidea, to the extent that Tassy⁴ has suggested that *B. grave* may represent the sister-group of other proboscideans. However, our knowledge of this species is insufficient to allow an accurate determination of its relationships with the form from Brezina. Several

authors^{6,7} have mentioned the occurrence of a small species of '*Barytherium*' in the late Eocene of Dor el Talha in Libya, known only by undescribed dental remains. A brief comparison indicates a size difference, the Libyan form being larger, and a few unimportant morphological differences in cheek-teeth morphology. On the basis of the very incomplete available data an ancestor-descendant relationship cannot be excluded.

The new Algerian form exhibits some derived characters shared in particular with the modern Proboscidea, that is, the Elephantidae and the Deinotheriidae, which suggest a close relationship. These are: (1) structure of the astragalus and calcaneum; (2) elevation and pneumatization of the skull; (3) ulna larger than radius; (4) radius fixed in a pronating position; (5) antero-posterior flattening and elongation of the femur.

However, the occurrence of a deep submaxillary fossa, which is an autapomorphic feature, rules out the possibility that this form is a direct ancestor of the Elephantidae-Deinotheriidae group. Nevertheless, it does appear as a potential sister-group of the modern Proboscidea. These new data demonstrate the great antiquity of proboscidean differentiation in Africa (Fig. 3).

The flora is represented by calcified charophyte gyrogonites, referred by M. Feist-Castel (USTL, Montpellier) to two species of the genus *Nitellopsis*: *N. (T.) dutemplei* (Watelet) and *N. (T.) aff. thaleri* (Castel and Grambast). The former is known from Sparnacian and Cuisian rocks of France, the latter from Cuisian and Lutetian beds of southern France and Spain. On the basis of these palaeobotanical data, a Cuisian, that is, late early Eocene age, can be suggested.

Received 24 October 1983; accepted 19 January 1984.

1. Andrew, C. W. *Tageblatt des V. Intern. Zoologischen-Kongress Berlin*, 6, 4-5 (1901).
2. Andrews, C. W. *A Descriptive Catalogue of the Tertiary Vertebrata of the Fayum, Egypt* (British Museum (Natural History), London, 1906).
3. Maglio, V. J. & Cooke, H. B. S. *Evolution of African Mammals* (Harvard University Press, 1978).
4. Tassy, P. *Géobios, Mém. spéc.* 6, 225-245 (1982).
5. Tassy, P. *Bull. Mus. natn. Hist. nat. Paris*, 4e sér. 3 C, 87-147 (1981).
6. Savage, R. J. G. *Symp. Geol. Libya*, 216-225 (University of Libya, Tripoli, 1971).
7. Harris, J. M. in *Evolution of African Mammals* (eds Maglio, V. J. & Cooke, H. B. S.) 315-332 (Harvard University Press, 1978).

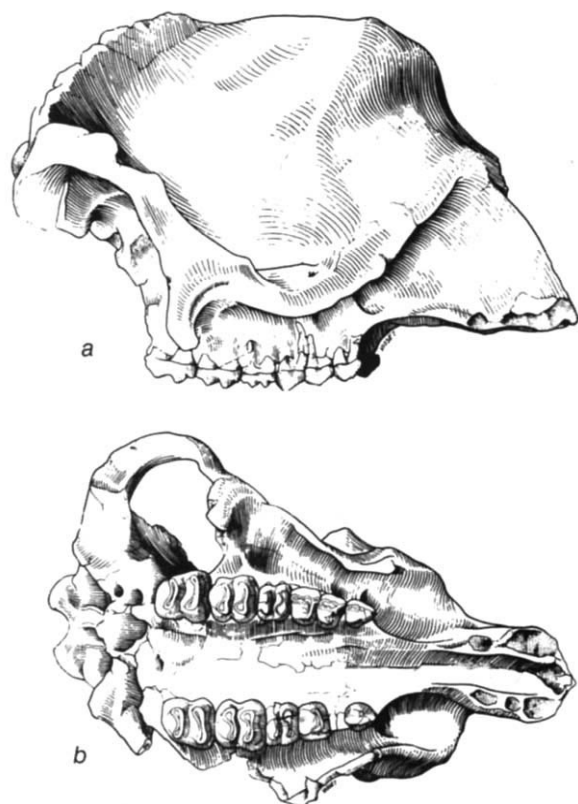


Fig. 2 a, Skull in lateral view. b, Skull in ventral view (drawings by D. Visset).

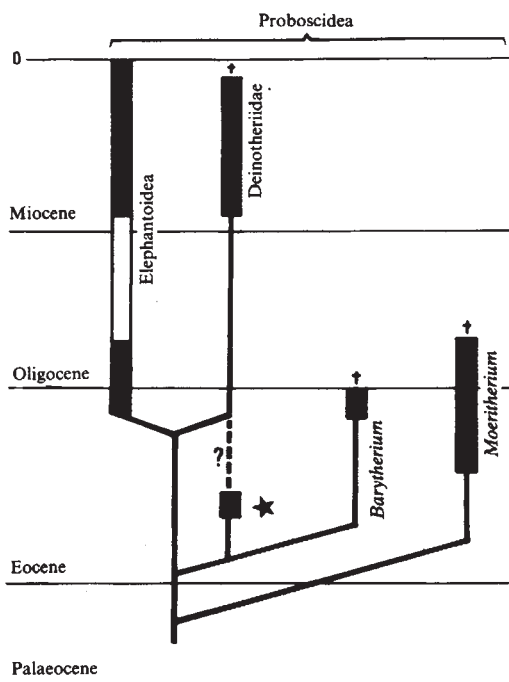


Fig. 3 Supposed phylogenetic relationships between the groups of proboscideans. Star represents the new proboscidean from El Kohol.

Experimental metastasis correlates with cyclic AMP accumulation in B16 melanoma clones

J. R. Sheppard*, Thomas P. Koestler, Steven P. Corwin, Charles Buscarino, John Doll, Bruce Lester, Russell G. Greig & George Poste

Department of Tumor Biology, Smith, Kline & French Laboratories, 1500 Spring Garden Street, Philadelphia, Pennsylvania 19101, USA

Metastasis is a complex process whereby tumour cells from a primary neoplastic growth disseminate throughout the body and establish secondary tumour foci in distant organs. Biochemical traits associated with, or essential for, the expression of the metastatic phenotype have not yet been identified^{1,2}. In the course of examining stimulation of the B16 murine melanoma adenylate cyclase by melanocyte-stimulating hormone (MSH) and by the diterpene forskolin, we noted that tumour cell clones³⁻⁵ isolated from common parent cell populations differed widely in their responses to these agonists. We report here that the accumulation of cyclic AMP induced by MSH or forskolin shows a strong positive correlation with the ability of B16 melanoma clones to form pulmonary tumour colonies when injected intravenously (i.v.) into syngeneic mice ('experimental

* Permanent address: Dight Institute of Human Genetics, 400 Church Street SE, University of Minnesota, Minneapolis, Minnesota 55455, USA.

metastasis'). In parallel *in vitro* analyses of cyclic AMP metabolism and *in vivo* assays of experimental metastasis using replicate cell preparations, highly metastatic tumour cell clones consistently show greater than a 30-fold increase in cellular cyclic AMP when exposed to MSH or forskolin. By contrast, clones with limited metastatic abilities respond to the same agonists with only a two- to threefold increase in cellular cyclic AMP. These data suggest that cyclic AMP metabolism is linked with biochemical pathways that are responsible for the formation of experimental metastasis by the B16 melanoma.

B16 clones were isolated and cultured *in vitro* as described previously⁶⁻⁸. Challenge with MSH or forskolin and measurement of cellular cyclic AMP was performed as described in Table 1 legend. MSH binds to specific cell-surface receptors that are coupled via a regulatory protein(s) to adenylate cyclase⁹. In contrast, forskolin interacts with the catalytic subunit of the adenylate cyclase^{10,11}, although it may also have other sites of action¹². In both cases, however, adenylate cyclase is activated and cyclic AMP formation is initiated. Both MSH and forskolin stimulate cyclic AMP accumulation in B16 melanoma clones but the magnitude of the response varied significantly among the different clones (Table 1). For example, clones F10-C52, F10-C1 and F1-C16 respond to MSH (1 μ M) challenge with 78-, 49- and 3-fold increases in cellular cyclic AMP levels respectively. Similar clonal heterogeneity was detected in response to forskolin challenge (Table 1). An investigation of the kinetics of cyclic AMP accumulation indicated that the limited response of certain clones (for example F1-C14) to either MSH or forskolin could not be attributed to a lag in cyclic AMP formation since in each case cyclic AMP accumulation in low-responsive clones attained a maximum value 10–30 min following challenge (Fig. 1).

To assess metastatic potential, replicate cultures of each B16 clone were injected *i.v.* into mice and 3 weeks later the number of pulmonary tumour colonies was estimated. This is referred to as experimental metastasis and although it does not encompass all the stages of the metastatic cascade⁶ the assay does provide a relatively rapid assessment of metastatic capacity that has been shown for the B16 melanoma to correlate well with the ability of B16 tumour cell variants to metastasize from a subcutaneous site ('spontaneous metastasis')⁷. It should be emphasized that for reasons discussed above and elsewhere^{8,13} the biological (experimental metastasis) and biochemical (hormonal challenge) assays were performed on the same day. Three B16 clones that responded minimally to MSH or forskolin displayed low metastatic potential. In striking contrast, four B16 clones that accumulated much higher levels of cellular cyclic AMP formed multiple pulmonary colonies, indicating that these clones had a high metastatic capacity (Table 1). In total, over 30 B16 clones of differing metastatic potential have been studied (not all data are shown here) and in each case we have found that their response to MSH or forskolin challenge *in vitro* is predictive of their metastatic capacity *in vivo*. Additional experiments have shown that the differences in metastatic potential are not related to growth rate *in vivo*, cell size or degree of pigmentation (data not shown). Although these data indicate a correlation between cyclic AMP metabolism and experimental metastasis it is unclear whether this is a causal relationship. Experiments to test this hypothesis and to extend these data to spontaneous metastasis are now in progress.

The potential value of cyclic AMP accumulation *in vitro* as a predictive marker for B16 metastatic behaviour of the clones is illustrated by other experiments in which clones that were presumed to be of low metastatic potential (based on earlier *in vivo* assays) were found to exhibit responses to MSH or forskolin that were contrary to the correlation defined above. However, retesting the clones *in vivo* revealed that their metastatic properties had increased significantly from the time they were last assayed (Table 2). Responsiveness of B16 melanoma clones to forskolin thus appears to offer an *in vitro* assay for predicting metastatic behaviour *in vivo* and for detecting drift in the

Table 1 Correlation between cyclic AMP accumulation in response to forskolin and MSH challenge and formation of experimental metastasis by B16 melanoma clones

B16 clone	Fold stimulation of basal cellular cyclic AMP		Median no. of experimental metastases (range)
	50 μ M forskolin (no. expts)	1 μ M MSH	
F10-C1	32 $\pm$ 4 (5)	49 $\pm$ 9 (3)	>300
F1-C5	31 $\pm$ 15 (3)	73 $\pm$ 14 (2)	130 (124–150)
F10-C23	35 $\pm$ 3 (6)	55 $\pm$ 6 (6)	177 (165–180)
F10-C52	36 $\pm$ 8 (2)	78 $\pm$ 10 (2)	>300
F1-C14	2 $\pm$ 1 (6)*	3 $\pm$ 1 (3)	5 (0–8)
F1-C16	3 $\pm$ 1 (4)*	3 $\pm$ 1 (4)	0 (0–3)
F1-C29	3 $\pm$ 1 (2)*	2 $\pm$ 1 (3)	2 (1–2)

Confluent cultures of B16 melanoma clones were collected with trypsin (0.25%) and EDTA (0.25%) as described elsewhere⁸. Cells were seeded into 24-well plates at 10⁵ cells per well in 1.0 ml of Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. Following a 18–24 h incubation in a humidified chamber containing 5% CO₂ the cultures were rinsed twice with 0.5 ml Dulbecco's phosphate-buffered saline (PBS). Forskolin (50 μ M) or MSH (1 μ M) were added in PBS and the cells incubated at 37 °C. These culture conditions are optimal and must be followed exactly in order to maximize the differences in hormonal responsiveness displayed by the B16 clones. The culture medium was aspirated and the cells treated with cold (4 °C) HClO₄ (0.7 M). This solution was neutralized with K₂CO₃ (0.35 M) and centrifuged (2,000 r.p.m. for 2 min) to remove precipitate. The cyclic AMP in the supernatant was measured by radioimmunoassay as previously described¹⁴. Cellular protein was determined by the method of Lowry *et al.*¹⁵. The average fold ($\pm$ s.e.m.) stimulation of basal cyclic AMP level following exposure to forskolin (10 min) or MSH (30 min) is shown. The basal level of cyclic AMP was approximately 18 pmol per mg protein for each B16 clone tested. Forskolin was purchased from Calbiochem (Irvine, California). Formation of experimental metastasis by the same clones was determined by injecting replicate cultures collected on the same day into C57BL/6 mice (10⁵ cells *i.v.*). The number of lung colonies was determined 18–21 days later. The data are expressed as the median value and range of lung colonies observed in at least five mice.

* To eliminate the possibility that certain B16 melanoma clones (such as F1-C14, F1-C16 and F1-C29) which, when challenged with forskolin (50 μ M) displayed only marginal stimulation of cyclic AMP accumulation, possessed a functional adenylate cyclase, cultures of these clones or membrane preparations derived from them were challenged with higher concentrations of forskolin (100–300 μ M). In these modified conditions, significant production of cyclic AMP was observed, indicating that these clones did possess an activatable adenylate cyclase. Moreover, inclusion of the phosphodiesterase inhibitor Ro 20-1724 (10 μ M) dramatically enhanced the accumulation of cyclic AMP when these same clones were challenged with forskolin (50 μ M) (data not shown).

metastatic phenotype during serial passaging of clones in culture⁴⁻⁶.

Cyclic AMP influences a variety of cellular functions¹⁴ and its role in regulating cell growth and the expression of various cellular characteristics has attracted considerable interest. Of particular relevance to the present study is evidence showing that Rous sarcoma virus-transformed avian cells exhibit exaggerated responses to a variety of agonists (prostaglandin E₁, purinergic and β -adrenergic) that mediate their effects by stimulating cellular synthesis of cyclic AMP^{15,16}.

Niles and Makarski¹⁷, in studies using polyclonal B16 melanoma lines of low (B16-F1) and high (B16-F10) metastatic abilities, found that high metastatic potential was associated with diminished responsiveness to MSH *in vitro*¹⁷. This is the reverse of the response observed in the present study. But in a subsequent study using clonal cultures, the same laboratory reported that, in contradiction to their earlier results, high metastatic potential was correlated with enhanced responsiveness to MSH¹⁸, a result consistent with the data presented here. Since heterogeneous polyclonal B16 cell lines were used in the initial studies by Niles and Makarski, changes in the balance of cellular subpopulations with high and low metastatic properties in the overall population could have occurred and thus account

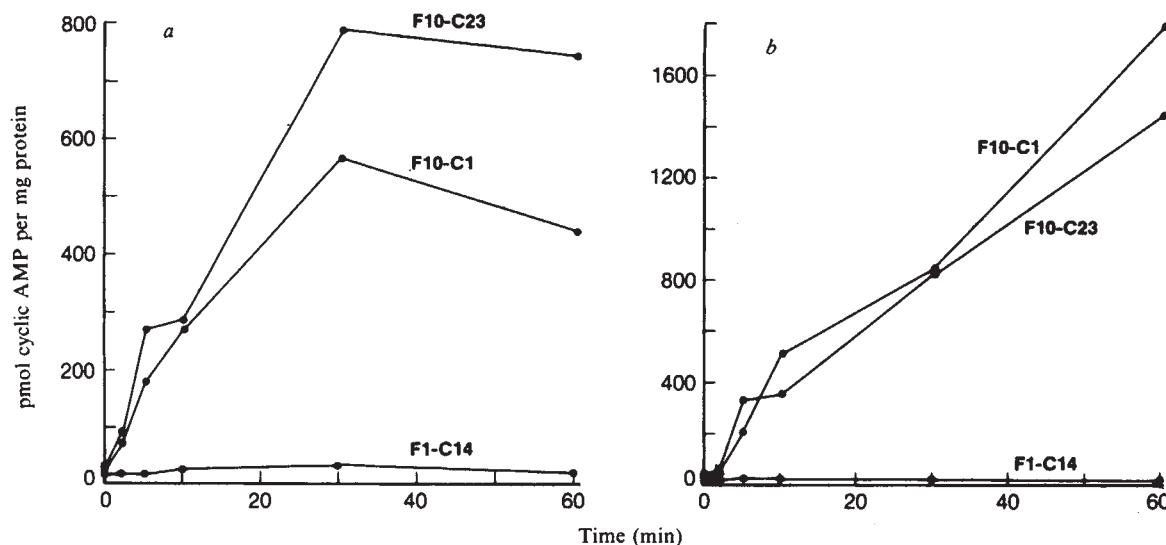


Fig. 1 Kinetics of cyclic AMP formation in B16 clones challenged with *a*, 1 μM MSH and *b*, 50 μM forskolin. Replicate cultures of each B16 clone were exposed to either MSH or forskolin for the time periods indicated and cyclic AMP assayed as described in Table 1 legend. Note that the ordinate scales in *a* and *b* are different.

Table 2 Correlation between instability of B16 metastatic phenotype in prolonged culture and responsiveness to forskolin challenge

B16 clone	Passage no.	Response to forskolin	Median no. of experimental metastases (range)
F1-C14	3	2 ± 1	5 (0-8)
	12	2 ± 1	4 (1-6)
	20	14 ± 5	51 (35-66)
	25	12 ± 3	66 (49-73)
	28	21 ± 6	76 (45-89)
F1-C16	3	3 ± 1	0 (0-3)
	22	20 ± 10	90 (68-102)

B16 clones were isolated and maintained *in vitro* as described in the text. Cultures were passed twice weekly and replicates tested at the indicated passage for experimental metastatic capacity *in vivo* (as described in Table 1 legend) and for forskolin responsiveness *in vitro* (the fold increase over the basal cellular cyclic AMP [$\pm$ s.e.m.] after exposure to 50 μM forskolin for 10 min).

for their disparate results. The inconclusive nature of these findings^{17,18} emphasizes the value of using clonal cultures in which the metastatic phenotype is defined in any attempt to determine the biochemical correlates of metastatic behaviour.

The molecular mechanism(s) accounting for the B16 clonal variation in response to MSH or forskolin are unknown but two possibilities have been eliminated by the study reported here. Differences in the characteristics and regulation of the MSH receptor are unlikely because forskolin activates adenylate cyclase through a receptor-independent mechanism. However, the enzyme's catalytic capacity or its intrinsic regulation (for example by GTP or divalent cations) could be altered in the high- and low-metastatic clones. Extracellular cyclic AMP accumulation in the various clones following MSH or forskolin challenge reflected the same variation observed in the intracellular accumulation of cyclic AMP (data not shown) indicating that differences in cyclic AMP efflux cannot account for the clonal variation.

Metastasis is a highly selective process in which only very few of the cells that escape the primary tumour survive to form metastases¹⁹. Heightened responsiveness of tumour cells to extracellular signals that stimulate intracellular cyclic AMP levels may enhance the ability of these cells to survive the hostile

environment encountered during the formation of secondary tumours. Cellular responsiveness to a stimulus that activates cyclic nucleotide metabolism could provide a mechanism whereby tumour cells could implement protective biochemical processes that facilitate their survival in the face of a wide range of potentially injurious assaults in much the same way that 'heat-shock' proteins can be induced by a variety of physical and chemical insults²⁰. Intracellular parasites²¹ and infectious agents^{22,23} display protective responses in mammalian hosts by mechanisms that promote cyclic AMP synthesis. Moreover, the proposed role of cyclic AMP as a metabolic stress signal in prokaryotic²⁴ and eukaryotic²⁵ cells, where it directly participates in the control of gene expression²⁶, invites speculation that cyclic AMP might serve a similar function in the survival of metastatic tumour cells.

We thank Jo Ann Cavallaro and Marcy Hughes for assistance in typing the manuscript.

Note added in proof: To test the generality of these findings, we have recently completed a survey of a series of clones derived from the highly invasive BL6²⁸ (a subline of the B16 tumour) and a murine melanoma of more recent origin, the K1735²⁹. Data from the BL6 clones are consistent with the correlation reported above. However, although there was significant variation among the K1736 clones in their responsiveness to forskolin and MSH and in their ability to form pulmonary tumour nodules when injected *i.v.*, there was no detectable correlation between cyclic AMP accumulation and metastatic potential.

Received 6 October 1983; accepted 13 January 1984.

- Fidler, I. J. & Kripke, M. L. *Science* **197**, 893-895 (1977).
- Nowell, P. C. *Science* **194**, 23-28 (1976).
- Fidler, I. J. & Poste, G. in *Tumor Cell Heterogeneity* (eds Owens, A. P., Coffey, D. S. & Baylin, S. B.) 127-145 (Academic, New York, 1982).
- Poste, G., Doll, J. & Fidler, I. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6226-6230 (1981).
- Harris, J. F., Chambers, A. F., Hill, R. P. & Ling, V. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5547-5551 (1982).
- Poste, G. & Fidler, I. J. *Nature* **283**, 139-146 (1980).
- Poste, G. *et al. Cancer Res.* **42**, 4086-4089 (1982).
- Greig, R. G. *et al. Cancer Res.* **43**, 6057-6065 (1983).
- Pawelek, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **72**, 951-955 (1975).
- Seaman, K. B., Padgett, W. & Daly, J. W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3363-3367 (1981).
- Insel, P. A., Stengel, D., Ferry, N. & Hanoune, J. *J. biol. Chem.* **257**, 7485-7490 (1982).
- Jakobs, K. H., Aktories, K. & Schultz, G. *Nature* **303**, 177-178 (1983).
- Nicolson, G. L. & Poste, G. *Curr. Problems Cancer* **8**, 4-42 (1983).
- Kebabian, J. & Nathanson, J. H. *Cyclic Nucleotides: Physiology and Pharmacology* (Springer, Berlin, 1982).
- Howard, R. F., Glossmann, H., Friis, R. R. & Sheppard, J. R. *Expl. Cell Res.* **139**, 436-439 (1982).
- Howard, R. F., Friis, R., Sheppard, J. R. & Glossmann, H. *J. cell. Physiol.* **111**, 295-302 (1982).

17. Niles, R. M. & Makarski, J. S. *J. cell. Physiol.* **96**, 355-360 (1978).
18. Niles, R. M. & Logue, M. P. *J. supramolec. Struct.* **11**, 251-258 (1979).
19. Fidler, I. J. & Hart, I. R. *Science* **217**, 998-1003 (1982).
20. Levinson, W., Opperman, H. & Jackson, J. *Biochim. biophys. Acta* **606**, 170-180 (1980).
21. Sheppard, J. R. *et al.* in *The Red Cell* (ed. Brewer, G.) 491-502 (Liss, New York, 1981).
22. Lowrie, D. B., Jackett, B. S. & Ratcliffe, N. A. *Nature* **254**, 600-602 (1975).
23. Confer, D. L. & Eaton, J. W. *Science* **217**, 948-950 (1982).
24. Busby, S. J. W., Aiba, H. & deCrombrugge, B. *J. molec. Biol.* **154**, 211-227 (1982).
25. Devreotes, P. N. in *Development of Dictyostelium Discoideum* (ed. Loomis, W. F.) 117-168 (Academic, New York, 1982).
26. Mangiarotti, G., Ceccarelli, A. & Lodish, H. F. *Nature* **301**, 616-618 (1983).
27. Lowry, O. H. *et al.* *J. biol. Chem.* **193**, 265-275 (1951).
28. Hart, I. R. *Am. J. Path.* **97**, 587-600 (1979).
29. Fidler, I. J. *et al.* *J. nat. Cancer Inst.* **67**, 947-959 (1981).

Inhibition of antigen-induced T-cell clone proliferation by antigen-specific antibodies

G. Corradin & H. D. Engers*

Department of Biochemistry, University of Lausanne, and

* Department of Immunology, Swiss Institute for Experimental Cancer Research, 1066 Epalinges, Switzerland

Attempts to inhibit the recognition of soluble antigens by T lymphocytes using antibodies specific for the antigen in question have been uniformly unsuccessful, in contrast to the observed specific inhibition of antibody generation by B cells¹. One exception is the unique situation whereby anti-hapten antisera inhibit the T-cell proliferative responses observed when hapten-specific T lymphocytes^{2,3} or clones⁴ are cultured with hapten-derivatized cells or proteins. The inability to inhibit T-cell functions by antigen-specific antibodies has been interpreted in several ways: (1) T cells possess a different repertoire from B cells; (2) the antibodies tested recognize epitopes present on the native antigen, whereas T cells recognize non-native (processed) structures; (3) the antigenic determinant(s) recognized by T cells on the surface of antigen presenting cells are either not accessible to antibodies, or are present in low amounts³. The development of antigen-specific T-cell clones and monoclonal antibodies both specific for the same antigenic determinants now allows this question to be investigated definitively. Here, we report for the first time the specific inhibition of antigen-induced T-cell clone proliferation by a monoclonal antibody directed against the relevant soluble protein antigen.

The derivation of apo-cytochrome *c* specific murine T-cell clones has been described elsewhere⁵. Clone 2-16 is a Lyt-1⁺2⁻ T-cell clone of BALB/c origin which is stimulated to proliferate by a variety of beef apo-cytochrome *c* peptides (but not haem-containing peptides), including one consisting of amino acids 11-25 of the cytochrome *c* molecule. Detailed studies on the antigenic specificity of this clone indicate that the actual segment recognized is located at residues 22-25 of the cytochrome *c* molecule⁶. Clone 1-46 has a different antigen specificity from clone 2-16, since it is stimulated by haem-containing peptide 1-65 and apo-peptide 1-65, but the exact site recognized by this clone is unknown⁷. Clone 12-A3, derived from a SJL mouse, is specific for apo-peptide 1-65 but not for haem-containing peptide 1-65, and thus is thought to possess a similar specificity to clone 2-16.

A monoclonal antibody (SJL 2-4; IgG2b) was derived from SJL mice immunized with beef apo-cytochrome *c*. It binds to peptide 11-25 and 23-104 and thus is thought to recognize a determinant similar, if not identical, to that seen by clone 2-16. Therefore we examined the ability of this monoclonal antibody to inhibit the antigen-induced proliferation of the three clones described above. As shown in Table 1, SJL 2-4 monoclonal antibody inhibited the apo-cytochrome *c*-induced proliferation

Table 1 Inhibition of apo-cytochrome *c* induced T-cell clone proliferation by the addition of an anti-apo-cytochrome *c* monoclonal antibody

Antigen ($\mu\text{g ml}^{-1}$)	Monoclonal antibody ($\mu\text{g ml}^{-1}$)	T-cell proliferation (c.p.m. $\times 10^{-3}$)		
		Clone 2-16	Clone 1-46	Clone 12-A3
Expt 1				
—	—	0.4 $\pm$ 0.1	0.5 $\pm$ 0.1	0.3 $\pm$ 0.1
10	—	8.7 $\pm$ 0.3	5.0 $\pm$ 0.3	7.1 $\pm$ 1.9
10	30	1.5 $\pm$ 0.0 (83)	5.6 $\pm$ 0.2 (0)	5.7 $\pm$ 2.6 (20)
10	100	0.3 $\pm$ 0.1 (100)	3.8 $\pm$ 0.2 (24)	2.6 $\pm$ 0.3 (63)
10	300	0.2 $\pm$ 0.0 (100)	2.5 $\pm$ 0.4 (50)	0.3 $\pm$ 0.1 (100)
50	—	8.5 $\pm$ 0.8	5.5 $\pm$ 0.5	11.5 $\pm$ 0.7
50	30	4.8 $\pm$ 0.6 (44)	5.5 $\pm$ 0.5 (0)	8.1 $\pm$ 1.0 (30)
50	100	3.1 $\pm$ 0.9 (63)	5.0 $\pm$ 0.3 (9)	8.5 $\pm$ 0.1 (25)
50	300	0.3 $\pm$ 0.2 (100)	ND	0.7 $\pm$ 0.3 (99)
200	—	6.3 $\pm$ 0.4	6.1 $\pm$ 0.5	12.2 $\pm$ 0.6
200	30	4.2 $\pm$ 0.4 (33)	7.0 $\pm$ 0.4 (0)	9.8 $\pm$ 0.4 (20)
200	100	5.0 $\pm$ 0.6 (20)	8.7 $\pm$ 0.1 (0)	11.6 $\pm$ 0.8 (5)
200	300	0.5 $\pm$ 0.2 (95)	5.6 $\pm$ 0.1 (8)	1.8 $\pm$ 0.7 (90)
Expt 2				
—	—	0.7 $\pm$ 0.1	ND	0.4 $\pm$ 0.2
20	—	30.2 $\pm$ 4.4	ND	11.5 $\pm$ 0.7
20	100	2.5 $\pm$ 1.8 (92)	ND	2.6 $\pm$ 0.3 (80)
20	300	0.2 $\pm$ 0.1 (100)	ND	0.3 $\pm$ 0.1 (100)

Monoclonal antibodies (SJL 2-4; IgG2b) specific for apo-cytochrome *c* were purified by ammonium sulphate precipitation followed by affinity chromatography over an apo-cytochrome *c* immunoabsorbant column. BALB/c T-cell clones 2-16, 1-46 and SJL T-cell clone 12-A3 (10^4 cells per 0.2 ml) were cultured with 5×10^5 irradiated (2,000 rad) syngeneic spleen cells and the indicated concentrations of apo-cytochrome *c* antigen and monoclonal antibodies. Cultures were pulsed after 48 h with $1 \mu\text{Ci}$ ^3H -thymidine for 16 h. The results represent the mean $\pm$ s.d. for triplicate cultures. Values given in parentheses represent the per cent inhibition of antigen-specific proliferation as compared with cultures in the absence of monoclonal antibodies. ND, not determined.

of clones 2-16 and 12-A3 in a dose-dependent fashion. However, proliferation of clone 1-46 was not inhibited by the same concentrations of SJL 2-4 antibody. These results indicate that the inhibition is specific and not due to an inherent non-specific toxic effect of the antibody preparation, nor to inhibition by the mere presence of antigen-antibody complexes. These

Table 2 Non-related T-cell clone proliferation is not inhibited by anti-apo-cytochrome *c* monoclonal antibodies

T-cell clone cultured	Antigen	Monoclonal antibody SJL 2-4 ($\mu\text{g ml}^{-1}$)	T-cell proliferation (c.p.m. $\times 10^{-3}$)
Expt 1 (Anti-M-MuLV)			
Clone MCH-14	—	—	0.1 $\pm$ 0.1
—	MBL-2.9	—	0.2 $\pm$ 0.1
Clone MCH-14	MBL-2.9	—	3.0 $\pm$ 0.1
Clone MCH-14	MBL-2.9	300	3.4 $\pm$ 0.1
Expt 2 (Anti- <i>L. tropica</i>)			
—	<i>L. tropica</i>	—	2.3 $\pm$ 0.3
Clone JH27.6.1	<i>L. tropica</i>	—	34.7 $\pm$ 1.2
Clone JH27.6.1	<i>L. tropica</i>	—	28.3 $\pm$ 1.9
	+apo-beef		
Clone JH27.6.1	<i>L. tropica</i>	300	20.1 $\pm$ 2.0
Clone JH27.6.1	<i>L. tropica</i>	300	23.5 $\pm$ 2.4
	+apo-beef		

The proliferation assay was performed as for Table 1. Expt 1: MCH-14 is a Lyt-1⁺2⁺ Moloney murine leukaemia virus (M-MuLV) specific C57BL/6 cytolytic T-cell clone derived from a regressor mouse by micromanipulation techniques⁸. MBL-2.9 is a Moloney-MuLV-derived C57BL/6 lymphoma which is lysed specifically by clone MCH-14. It is irradiated (7,000 rads) when used as a source of antigen in cultures (3×10^4 per 0.2 ml), together with irradiated (2,000 rad) syngeneic spleen cells (1×10^6 per 0.2 ml). Expt 2: JH27.6.1 is a Lyt-1⁺2⁻ *Leishmania tropica* specific non-cytolytic CBA T-cell clone derived from *in vitro* cultured immune lymph node cells by micromanipulation techniques⁹. It proliferates specifically *in vitro* when stimulated with freeze-thawed *L. tropica* parasites (10^6 per 0.2 ml) as a source of antigen, together with 1×10^6 irradiated (2,000 rad) syngeneic spleen cells as a source of accessory cells. The concentration of apo-beef cytochrome *c* used was $200 \mu\text{g ml}^{-1}$.

results were confirmed by testing the effects of the monoclonal antibody or antigen-antibody complexes on the proliferative response of a Moloney leukaemia virus-specific Lyt-1^+2^+ cytolytic T-cell clone⁷ and a *Leishmania tropica* specific Lyt-1^+2^- T-cell clone⁸, neither of which was significantly inhibited (Table 2).

When irradiated spleen cells were 'pulsed' with beef apo-cytochrome *c* before subsequent culture with the T-cell clone 2-16 and the SJL 2-4 monoclonal antibody (Table 3), a similar dose-dependent inhibition of antigen-induced proliferation was observed. These results suggest that the inhibition was not due to an impairment of the uptake and/or appropriate processing of the apo-cytochrome *c* molecules. The SJL 2-4 antibody also inhibited the antigen-induced proliferation of two additional long-term T-cell lines and three clones (of both BALB/c and SJL origin) which are specific for the apo-1-38 fragment of the cytochrome *c* molecule (data not shown). Since the monoclonal antibody was generated by immunization of SJL mice, these results suggest that a similar cytochrome *c* peptide is generated during antigen processing in both SJL and BALB/c strains of mice.

Table 3 T-cell clone proliferation induced by 'antigen pulsed' spleen cells is inhibited by monoclonal antibodies directed against apo-cytochrome *c*

	Apo-cytochrome <i>c</i> used to pulse spleen cells ($\mu\text{g ml}^{-1}$)	Monoclonal antibodies ($\mu\text{g ml}^{-1}$)	Proliferation of clone 2-16 (c.p.m. $\times 10^{-3}$)
Expt 1	500	—	2.2 $\pm$ 0.2
		50	0.9 $\pm$ 0.1 (59)
		100	0.4 $\pm$ 0.1 (100)
		300	0.3 $\pm$ 0.1 (100)
	2,000	—	6.0 $\pm$ 1.0
		50	5.5 $\pm$ 0.2 (9)
Expt 2	500	100	6.5 $\pm$ 0.7 (0)
		300	0.8 $\pm$ 0.1 (87)
	500	—	5.9 $\pm$ 0.3
		300	0.8 $\pm$ 0.1 (93)

Irradiated (2,000 rad) BALB/c spleen cells (10^6) were incubated for 2 h with different concentrations of apo-cytochrome *c* and washed four times with medium before the addition of 10^4 T-cell clone 2-16. Purified SJL 2-4 monoclonal antibodies were used as described in Table 1 legend. The proliferation of control cultures in the absence of added apo-cytochrome *c* was 400 c.p.m. Values in parentheses represent the per cent inhibition of T-cell clone proliferation.

Thus, it should now be feasible to purify and elucidate the biochemical structure of the defined antigenic determinant presented on the macrophage cell surface which subsequently stimulates antigen-specific T-cell clones to proliferate and secrete lymphokines.

We thank C. Bron, J. Louis and J.-C. Cerottini for their help and discussions, A. Chambaz for technical assistance, and M.-C. Knecht for secretarial assistance. This work was supported by grants from the Swiss Science Foundation (3.643.-0.80) and the E. R. Shepard Foundation/K.F. Associates.

Received 12 September 1983; accepted 17 January 1984.

Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis

M. A. H. Surani, S. C. Barton & M. L. Norris

AFRC Institute of Animal Physiology, 307 Huntingdon Road, Cambridge CB3 0JQ, UK

It has been suggested that the failure of parthenogenetic mouse embryos to develop to term is primarily due to their aberrant cytoplasm and homozygosity leading to the expression of recessive lethal genes¹. The reported birth of homozygous gynogenetic (male pronucleus removed from egg after fertilization) mice and of animals following transplantation of nuclei from parthenogenetic embryos to enucleated fertilized eggs^{2,3}, is indicative of abnormal cytoplasm and not an abnormal genotype of the activated eggs. However, we⁴ and others^{5,6} have been unable to obtain such homozygous mice. We investigated this problem further by using reconstituted heterozygous eggs, with haploid parthenogenetic eggs as recipients for a male or female pronucleus. We report here that the eggs which receive a male pronucleus develop to term but those with two female pronuclei develop only poorly after implantation. Therefore, the cytoplasm of activated eggs is fully competent to support development to term but not if the genome is entirely of maternal origin. We propose that specific imprinting of the genome occurs during gametogenesis so that the presence of both a male and a female pronucleus is essential in an egg for full-term development. The paternal imprinting of the genome appears necessary for the normal development of the extraembryonic membranes and the trophoblast.

Eggs from (C57BL/6J $\times$ CBA/Ca)F₁ females (henceforth called F₁) were activated and the haploid parthenogenetic eggs selected at about 5 h post-activation to serve as recipient eggs for either a male or a female pronucleus taken from fertilized eggs (see Table 1). The donor male pronuclei were obtained from F₁ $\times$ MF1 δ eggs and the female pronuclei from MF1 δ $\times$ F₁ δ eggs or F₁ δ $\times$ MF1 δ eggs. Reconstituted eggs with two distinct pronuclei were cultured overnight, when all of them cleaved to the two-cell stage. These embryos were transferred to the right oviducts and control MF1 $\times$ MF1 embryos to the left oviducts of day 1 MF1 pseudopregnant mice (day 1 = day of vaginal plug).

The first series of experiments produced, at day 15, two normal embryos from five haploid eggs reconstituted with a male MF1 pronucleus. By contrast, in 29 reconstituted haploid eggs into which a second female MF1 pronucleus was introduced, 15 embryos implanted after transfer of the eggs but all the implantation sites were resorbing without any detectable embryonic derivatives. Therefore, all of the former type of reconstituted eggs were allowed to proceed to term (day 19–20 of pregnancy) while the rest were examined on day 10 of pregnancy to assess the cause of their failure to develop. Day 10 of pregnancy was chosen from our previous experience of development of gynogenetic⁴ and parthenogenetic⁷ embryos which deteriorate after this time.

The results show that introduction of a male MF1 pronucleus can rescue parthenogenetic haploid eggs, with nearly 38% of them reaching term but a female F₁ or MF1 pronucleus is unable to do so (see Table 1). However, between 50 and 79% of the latter implanted, which shows that most of them had at least developed to the blastocyst stage, but all the decidual swellings at day 10 were about half the volume of their normal controls whether or not they contained embryos. The few embryos present were small or retarded with poor extraembryonic membranes and trophoblast. The most advanced embryo was perfectly formed but it was very small; for a brief description of other embryos see Fig. 1 legend.

1. Uhr, J. & Moeller, G. *Adv. Immun.* **8**, 81–126 (1968).
2. Thomas, D. W. & Shevach, E. M. *J. Immun.* **121**, 1145–1151 (1978).
3. Schevach, E. M., Chan, C. & Clement, L. T. *Eur. J. Immun.* **12**, 819–824 (1982).
4. Clayberger, C., Dekruyff, R. H., Aisenberg, J. & Cantor, H. *J. exp. Med.* **157**, 1906–1919 (1983).
5. Corradin, G., Zubler, R. H. & Engers, H. D. *J. Immun.* **127**, 2442–2446 (1981).
6. Corradin, G., Juillerat, M. A., Vita, C. & Engers, H. D. *Molec. Immun.* **20**, 763–768 (1983).
7. Weiss, A., Brunner, K. T., MacDonald, H. R. & Cerottini, J.-C. *J. exp. Med.* **152**, 1210–1229 (1980).
8. Louis, J. A. *et al. Immun. Rev.* **61**, 215–243 (1982).

Table 1 Development of haploid activated eggs reconstituted with a male or a female pronucleus

Type of reconstituted egg*	No. of eggs		Embryos†				
	With fused karyoplasts/total operated	Transferred to recipients that became pregnant/total transferred	Implanted	Normal	Small	Retarded and abnormal	None
(1) Pn + Ak MF1	30/65	24/30	—	9 Live young‡			
(2) Pn + Gk MF1	57/96	48/57	38	0	8	12	18
(3) Pn + Gk F ₁	38/73	24/38	12	0	0	4	8
(4) Pn + Pnk	15/22	7/15	4	0	0	0	4
(5) Unoperated fertilized MF1 × MF1	—	106/162	88	65	4	2	17

Cumulus masses containing unfertilized eggs were recovered 16.5–17.0 h after an injection of 5 IU human chorionic gonadotropin (HCG) in animals which had been injected previously with 5 IU pregnant mare's serum (48 h before injection of HCG) to induce superovulation. The cumulus cells were dispersed by incubation in 300 IU ml⁻¹ hyaluronidase (ovine testis) in phosphate-buffered saline (PBI)¹⁹ plus 0.4% bovine serum albumin (BSA) for 3–5 min at 37 °C. Eggs were activated by culture at room temperature for 7 min in M16 plus 0.4% BSA plus 7% absolute ethanol²⁰, washed 6–9 times in M16 plus 0.4% BSA²¹ and cultured in this medium under paraffin oil at 37 °C in humidified 5% CO₂ in air. The second polar body was extruded and a single pronucleus was detected after ~5 h in 70–90% of the parthenogenetic activated eggs. Fertilized eggs were obtained at the same time from F₁♀ × MF1♂ and MF1♀ × F₁♂ matings. The transfer of donor pronuclei from fertilized eggs to haploid parthenogenetic eggs was carried out on a Leitz micromanipulator²² essentially as described previously²³. The donor and recipient eggs were placed in PBI + 0.1% BSA containing the cytoskeletal inhibitors cytochalasin D and Nocodazole (Sigma) at a final concentration of 1 µg ml⁻¹ and 0.05 µg ml⁻¹ respectively, dissolved in 1.5 µl dimethyl sulphoxide. After 30–45 min at 37 °C, the eggs were suspended in hanging drops for micromanipulation. Using a 25-µm bevelled needle, donor pronuclei were drawn into the needle and the membrane pinched off to make a small karyoplast fragment, followed by approximately 6 pl of β-propiolactone-inactivated Sendai virus^{24,25} (3,000 haemagglutination units ml⁻¹) to induce fusion²⁶. Virus and karyoplast were then gently released into the perivitelline space underneath the zona pellucida of the recipient haploid parthenogenetic eggs, usually at the opposite end to the polar body. The recipient eggs with adhering karyoplasts were washed twice in PBI plus 0.1% BSA and cultured in M16 plus 0.4% BSA at 37 °C as described above. The eggs were checked for fusion of the karyoplast ~4 h after the completion of manipulation on the last group of eggs. Fusion was confirmed by the disappearance of the karyoplast and the appearance of two pronuclei inside the recipient egg. The fusion rate was between 30 and 65%. The diploid reconstituted eggs were separated from the haploid ones with unfused karyoplasts, washed nine times in equilibrated M16 plus 0.4% BSA and cultured overnight in this medium at 37 °C in 5% CO₂ in air under paraffin oil. The following morning all the eggs had cleaved, and they were transferred to the oviducts of day 1 MF1 outbred albino pseudopregnant mice.

* (C57BL × CBA)F₁ haploid activated eggs: Pn (parthenogenone) were used as recipients for a male pronucleus, Ak MF1, from F₁♀ × MF1♂ eggs (1). The haploid eggs were also used as recipients for a female pronucleus, Gk MF1, from MF1♀ × F₁♂ eggs (2), or Gk F₁, from F₁♀ × MF1♂ eggs (3). In a few haploid eggs, a second pronucleus from activated eggs, Pnk, was introduced (4).

† The F₁ animals are non-albino, homozygous for glucose phosphate isomerase (GPI *bb*) while the MF1 are outbred albino (GPI *aa*). The live young and the other experimental and control embryos were typed for GPI²⁷. The live young (1) and the experimental embryos in (2) were GPI *ab*, the experimental embryos in (3) were GPI *bb* and the controls (5) were GPI *aa*. All the embryos were examined on day 10 of pregnancy apart from those in group (1) which were allowed to go to term.

‡ Five females and four males were obtained. All the animals have now reached adulthood and they have bred normally.

Carry-over of a small amount of fertilized egg cytoplasm (2–5% of the total egg volume) and plasma membrane cannot account for these results as it occurred to an equal degree in both male and female pronuclei. It is unlikely that a specific cytoplasmic region from fertilized eggs was included with the male pronucleus because the site of sperm entry into the egg is random and the pronuclei migrate continuously from the periphery to the centre of the egg. We cannot exclude entirely the possible transfer of specific extragenetic components closely associated with the male pronucleus. However, when the eggs are introduced into the cytoskeletal inhibitors before micromanipulation (see Table 1), they lose their rigidity, becoming more 'fluid' so that considerable mixing of the cytoplasmic contents occurs during the extraction of the donor pronuclei. Furthermore, we have not obtained development of activated eggs to term by introducing cytoplasm from fertilized eggs (unpublished observations), nor of diploid gynogenones⁴ and, from preliminary results, of heterozygous gynogenetic embryos. Taken together, the results on the rescue of parthenogenones by a male pronucleus do not support the notion that the cytoplasm of activated eggs is aberrant^{2,3}. The failure of development to term of heterozygous reconstituted eggs having two female pronuclei also does not support the contention that homozygosity is a primary cause of the death of gynogenones⁴ and parthenogenones^{1,6,7,8}. Although a male and a female pronucleus are presumably equivalent in their genetic contribution to the embryo, the crucial difference lies in their origin as the former is derived after spermatogenesis and the latter after oogenesis. We propose that specific imprinting of the paternal and maternal genomes occurs during gametogenesis, which is perhaps why both of them are required in an egg for full-term development.

There is good evidence for paternal imprinting of the mouse genome and some paternal genes are expressed as early as the two-cell stage⁹. In the extraembryonic membranes and trophoblast of female embryos, the paternal X chromosome is preferentially inactivated^{10–12}. There are conflicting results on X-inactivation in these tissues in parthenogenetic embryos^{13,14}, but variable X-inactivation has been suggested as a cause of their poor development¹³. In addition, studies on DDK mutant mice suggest that the cytoplasm of eggs from this strain of mice cannot activate some specific paternal genes from alien males, essential for the development of embryos after implantation and especially for the normal development of the trophoblast¹⁵. On the basis of these results, reconstituted eggs having two female pronuclei and lacking a paternal genome are likely to be affected in a similar way, which they apparently are.

If the primary defect in embryos having the maternal genotype is in the extraembryonic membranes, this may explain the apparent discrepancy between the ability of these embryonic cells to proliferate and differentiate in ectopic sites^{16,17} and chimaeras^{8,18}, and their poor development when relying on their own extraembryonic tissues to sustain them. Development of embryos with two female pronuclei in this study is no different from that of gynogenetic⁴ or parthenogenetic embryos⁷. Nearly 25% of the parthenogenetic embryos can reach about this level of development but deteriorate rapidly thereafter (ref. 7 and S.C.B. and M.A.H.S., unpublished). This is perhaps not surprising as the embryo increasingly relies on the extraembryonic tissue for nutrition, which develops particularly poorly in the absence of the paternal genome. However, there could also be other less obvious effects of the genetic constitution of eggs on embryonic development.

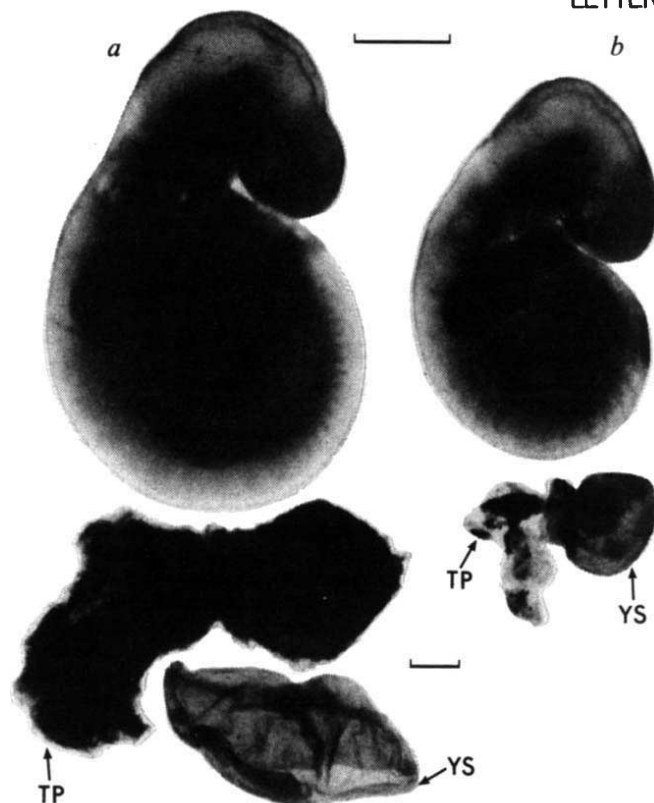


Fig. 1 Two embryos obtained on day 10 of pregnancy (scale bar, 0.5 mm): *a*, control MF1 x MF1 embryo (GPI *aa*); *b*, the most advanced embryo obtained from a reconstituted parthenogenetic haploid egg with an MF1 female pronucleus (GPI *ab*). Their corresponding trophoblasts (TP, mural trophoblast and ectoplacental cone) and yolk sacs (YS) are shown directly underneath the embryos at a reduced magnification (scale bar, 0.5 mm). The experimental embryo, apart from its small size, was apparently normal, with about 25 somites and a beating heart. Note the greatly reduced trophoblast tissue obtained with the experimental embryo. Nine live-born non-albino mice (GPI *ab*) were obtained from reconstituted haploid parthenogenetic F_1 eggs (GPI *b*) with an MF1 male pronucleus (GPI *a*) (see Table 1). All the embryos resulting from the reconstituted eggs with a female F_1 or MF1 pronucleus were cramped inside tiny yolk sacs and surrounded by very sparse trophoblast. There were two 20–25-somite embryos, three 15-somite embryos and three 6–8-somite embryos. All the embryos were less than half the volume of their normal controls of comparable somite stage, and all in very deficient membranes. There were two very abnormal embryos (one tail part only, one contorted deteriorating neural tissue), and a further 14 tiny yolk sac vesicles, often with a prominent Reichert's membrane. The remaining 26 implantation sites contained no embryonic derivatives.

Our results are in contrast to the reported birth of gynogenetic and parthenogenetic mice^{2,3}. One explanation could be that in rare cases enough trophoblast develops to carry the embryos to term. Incomplete enucleation of some eggs also apparently occurred in previous studies^{2,3} but whether this permits sufficient participation by the paternal genome to ensure development to term is unknown. Nevertheless, our studies suggest that specific imprinting of the genome during gametogenesis is essential for full-term development in the mouse. At some point during embryonic development and formation of primordial germ cells, previous influences on the genome are presumably lost and new ones initiated during the production of gametes.²⁸

We thank Catharine Adams and Dr B. Mahy for their considerable help during this study.

Received 5 December 1983; accepted 27 February 1984.

1. Graham, C. F. *Biol. Rev.* **49**, 399–422 (1974).
2. Hoppe, P. C. & Illmensee, K. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5657–5661 (1977).
3. Hoppe, P. C. & Illmensee, K. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1912–1916 (1982).

4. Surani, M. A. H. & Barton, S. C. *Science* **222**, 1034–1036 (1983).
5. Modlinski, J. A. *J. Embryol. exp. Morph.* **60**, 153–161 (1980).
6. Markert, C. L. *J. Hered.* **73**, 390–397 (1982).
7. Kaufman, M. H., Barton, S. C. & Surani, M. A. H. *Nature* **265**, 53–55 (1977).
8. Surani, M. A. H., Barton, S. C. & Kaufman, M. H. *Nature* **270**, 601–603 (1977).
9. Sawicki, J. A., Magnuson, T. & Epstein, C. J. *Nature* **294**, 450–451 (1981).
10. West, J. D., Frels, W. I., Chapman, V. M. & Papaioannou, V. E. *Cell* **12**, 873–882 (1977).
11. Takagi, N., Wake, N. & Sasaki, M. *Cytogenet. Cell Genet.* **20**, 240–248 (1978).
12. Harper, M. I., Fosten, M. & Monk, M. *J. Embryol. exp. Morph.* **67**, 127–135 (1982).
13. Endo, S. & Takagi, N. *Jap. J. Genet.* **56**, 349–356 (1981).
14. Rastan, S., Kaufman, M. H., Handyside, A. H. & Lyon, M. F. *Nature* **288**, 172–173 (1980).
15. Wakasugi, N. *J. Reprod. Fert.* **41**, 85–96 (1974).
16. Stevens, L. C. *Symp. Soc. dev. Biol.* **33**, 93–106 (1975).
17. Iles, S. A., McBurney, M. W., Bramwell, S. R., Deussen, Z. A. & Graham, C. F. *J. Embryol. exp. Morph.* **34**, 387–405 (1975).
18. Stevens, L. C., Varnum, D. S. & Eicher, E. M. *Nature* **269**, 515–517 (1977).
19. Whittingham, D. G. & Wales, R. G. *Aust. J. biol. Sci.* **22**, 1065–1072 (1969).
20. Cuthbertson, K. S. R. *J. exp. Zool.* **226**, 311–314 (1983).
21. Whittingham, D. G. *J. Reprod. Fert. Suppl.* **14**, 7–21 (1971).
22. Barton, S. C. & Surani, M. A. H. *Expl. Cell Res.* **146**, 187–191 (1983).
23. McGrath, J. & Solter, D. *Science* **220**, 1300–1302 (1983).
24. Neff, J. M. & Enders, J. F. *Proc. Soc. exp. Biol. Med.* **127**, 260–271 (1968).
25. Giles, R. E. & Ruddle, F. H. *In Vitro* **9**, 103–108 (1973).
26. Graham, C. F. *Acta endocr. Suppl.* **153**, 154–167 (1971).
27. Chapman, V. M., Whitten, W. K. & Ruddle, F. H. *Dev. Biol.* **26**, 153–161 (1971).
28. Markert, C. L. & Seidel, G. E. in *New Technologies in Animal Breeding* (eds Brackett, B. G., Seidel, G. E. & Seidel, S. M.) 181–199 (Academic, New York, 1981).

Maternal T^{hp} lethality in the mouse is a nuclear, not cytoplasmic, defect

James McGrath & Davor Solter

The Wistar Institute of Anatomy and Biology, 36th Street at Spruce, Philadelphia, Pennsylvania 19104, USA

The T^{hp} mutation, an allele of the mouse T/t complex, differs from other known mutations in that its effects are determined by the sex of the parent from which it is inherited; when inherited from the female parent, it is invariably lethal at the embryonic stage, but, most embryos which inherit the mutation from the male parent survive^{1,2}. Thus most heterozygous embryos carrying the maternally derived mutation die in the second half of pregnancy, while the exceptional embryos surviving to parturition give oedematous, cyanotic individuals that die within 24 h^{3,4}. The lethal maternal effect of T^{hp} may be transmitted either through the cytoplasm of the ovum (oogenic defect) or through the female pronucleus (embryogenic defect). Here we have sought to decide between these possibilities by performing reciprocal nuclear transplantations⁵ between one-cell embryos from $T^{hp}/+$ and $+/+$ females. Our observation that this maternally inherited lethal effect of T^{hp} persists when $T^{hp}/+$ pronuclei are transplanted into $+/+$ cytoplasm suggests that the defect responsible for the pattern of inheritance lies in the pronuclei and not the cytoplasm.

One cell-stage embryos obtained from $+/+$ albino females mated to $+/+$ albino males were used in reciprocal nuclear transplantations with one-cell stage embryos obtained from pigmented $T^{hp}/+$ females previously mated to pigmented $+/+$ males. Nuclear transplant embryos were transferred to the oviducts of day 1 pseudo-pregnant females and allowed to develop to term. The number, coat colour phenotype and tail length of the progeny were observed, as T^{hp} causes a shortening of the tail when heterozygous.

Of 197 successful nuclear transplant embryos (Table 1) in which wild-type pronuclei from albino females mated to albino males were introduced into the cytoplasm of enucleated embryos from pigmented $T^{hp}/+$ females, 45 (20 females and 25 males) normal-tailed albino progeny resulted (23%). Among 206 successful nuclear transplant embryos in which pronuclei from pigmented $T^{hp}/+$ females mated to pigmented males were introduced into the cytoplasm of enucleated embryos from albino females, 18 (10 females and 8 males) pigmented progeny were born (9%). Thus, a significantly greater proportion of nuclear transplant embryos, in which the egg cytoplasm was derived primarily from $T^{hp}/+$ females, developed to term (23%) than those embryos in which egg cytoplasm was derived primarily from $+/+$ females (9%) ($\chi^2=8.99$; $P<0.01$). We

therefore find no evidence that ovum cytoplasm from $T^{hp}/+$ females inhibits normal embryogenesis.

Since the $T^{hp}/+$ females were mated to $+/+$ males, half of the nuclear transplant embryos should be $T^{hp}/+$ (short-tailed) and half $+/+$ (normal-tailed). Among the 18 progeny which resulted from the introduction of pronuclei from $T^{hp}/+$ females into $+/+$ cytoplasm (Table 1), a significantly greater number of normal-tailed progeny (16) than short-tailed progeny (2) ($\chi^2=9.38$; $P<0.01$) were born. Of the 16 normal-tailed progeny, 2 were dead at birth while the remaining 14 were viable and survived to adulthood. Of the two $T^{hp}/+$ progeny, one was dead at birth while the other was oedematous, cyanotic and died within 24 h of parturition. The number of viable normal-tailed progeny (14) is significantly greater than the number of viable short-tailed progeny (0) ($\chi^2=12.06$; $P<0.001$). Because the introduction of $T^{hp}/+$ pronuclei into wild-type cytoplasm did not result in the birth of viable $T^{hp}/+$ progeny we conclude that T^{hp} lethality is inherited as a nuclear, not cytoplasmic, defect.

The present evidence, which indicates a nuclear origin of the T^{hp} lesion, is consistent with previous models of T^{hp} lethality that have proposed differential functioning of the maternal and paternal genome in the proximal region of chromosome 17 during normal embryogenesis. Johnson has suggested that there may be an 'activator' locus on the maternal chromosome 17 that is required for proper gene expression at an unidentified second locus³. McLaren has suggested that the maternal genome may be normally active at the T^{hp} chromosomal region while its paternal counterpart is preferentially inactivated⁶. This inactivation, it is suggested, might occur in a manner akin to the preferential paternal X-chromosome inactivation that occurs in murine extra-embryonic tissues. Finally, in a study of translocation heterozygotes in which embryos could inherit both proximal portions of chromosome 17 solely from the male parent, or solely from the female parent, Lyon and Glenister noted a decrease in the number of progeny inheriting both proximal portions of chromosome 17 from the male parent⁷. The authors suggested that, as with T^{hp} , a functional maternal chromosome 17 may be required for successful embryogenesis.

T^{hp} is believed to be a deletion of a proximal portion of chromosome 17 extending distally beyond *qk* and including the *Tcp-1* locus^{8,9}. The fact that a female embryo inheriting the T^{hp} -deleted chromosome from the male parent is viable but her offspring that receive the T^{hp} -deleted chromosome are not,

strongly suggests differential modification of this region of chromosome 17 in males and females. Presumably this modification is required for differential expression of the maternal and paternal genome in the T^{hp} region in the subsequent embryonic generation. Of interest is the time during embryogenesis when this differential expression of the maternal and paternal chromosomes occurs and whether it occurs in all or a few embryonic or extra-embryonic tissues.

This work was supported by NCI grants CA-10815 and CA-25875 and by grants HD-12487 and HD-17720 from the National Institute of Child Health and Human Development. J. McG. was supported by NCI grant T32-CA-09171.

Received 18 November 1983; accepted 17 January 1984.

1. Dickie, M. M. *Mouse Newsl.* **32**, 43-44 (1965).
2. Johnson, D. R. *Genetics* **76**, 795-805 (1974).
3. Johnson, D. R. *Genet. Res.* **24**, 207-213 (1975).
4. Alton, A. thesis, Cornell Univ. (1982).
5. McGrath, J. & Solter, D. *Science* **220**, 1300-1302 (1983).
6. McLaren, A. in *Maternal Effects in Development* (eds Newth, D. R. & Balls, M.) 287-320 (Cambridge University Press, 1979).
7. Lyon, M. F. & Glenister, P. H. *Genet. Res.* **29**, 83-92 (1977).
8. Bennett, D. et al. *Genet. Res.* **26**, 95-108 (1975).
9. Silver, L. M., Artzt, K. & Bennett, D. *Cell* **17**, 275-284 (1979).
10. Whitten, W. K. *Adv. Biosci.* **6**, 129-139 (1971).
11. Abramczuk, J., Solter, D. & Koprowski, H. *Dev. Biol.* **61**, 378-383 (1977).

Nucleotide sequence of mutant $I-A\beta^{bm12}$ gene is evidence for genetic exchange between mouse immune response genes

Katherine R. McIntyre & J. G. Seidman

Department of Genetics, Harvard Medical School, 45 Shattuck Street, Boston, Massachusetts 02115, USA

Immune response genes^{1,2} of the murine major histocompatibility complex encode cell-surface glycoproteins that are expressed predominantly on B cells and macrophages and regulate immune responsiveness by restricting antigen recognition by T cells. The two classes of immune response molecule, termed I-A and I-E, are each comprised of two polymorphic chains (α and β), and nucleotide sequence analysis of genomic or cDNA clones has revealed that most of the amino acid differences between allelic I-A α or β chains occur in the first extracellular domain^{3,4}. The mutant mouse strain B6.C-H-2^{bm12} (bm12), which differs from its parental strain C57BL/6 (B6) at the $I-A\beta$ locus⁵⁻¹⁴, exhibits an immune response profile markedly different from that of B6. Here we present the nucleotide sequence of the mutant bm12 $I-A\beta$ gene. Sequence comparison within the coding regions reveals three productive nucleotide differences between the $I-A\beta$ genes of B6 and bm12 mice, all three differences occurring within a stretch of 14 nucleotides in the exon encoding the first extracellular domain. The clustered nature of the bm12 mutation, as well as the specific amino acid changes it engenders, suggest a possible mechanism for the generation of polymorphism in class II antigens.

The structure of the $I-A\beta^b$ gene has been completely determined^{4,15} and is shown schematically in Fig. 1. A 6.5-kilobase (kb) *EcoRI* fragment carrying most of the $I-A\beta^{bm12}$ gene was cloned from bm12 genomic DNA into bacteriophage λ and subsequently into the plasmid pBR327 (see Fig. 1). Single and double digests of plasmid subclones as well as genomic DNA, using a number of different restriction enzymes (including *BamHI*, *EcoRI*, *HindIII*, *SmaI* and *XhoI*), and Southern blot analysis with probes spanning the β_1 , β_2 , TM and IC₁ exons revealed no differences between the $I-A\beta^b$ and $I-A\beta^{bm12}$ alleles (data not shown). This result indicates that the mutation which has altered the primary structure of the $I-A\beta^{bm12}$ polypeptide does not involve any gross insertions or deletions in either the exons or the introns of this gene.

Table 1 Phenotype of nuclear transplant offspring

Nuclear genotype	Cytoplasmic genotype	No. of successful nuclear transplants	No. of progeny Normal tail	Short tail
$T^{hp}/+$, $+/+$	$+/+$	206	16	2
$+/+$	$T^{hp}/+$, $+/+$	197	45	0

Breeding stock of $T^{hp}/+$ males and females were kindly provided by Drs D. Bennett and H. Axelrod. $T^{hp}/+$ embryos were obtained from brown or agouti $T^{hp}/+$ females mated to C57BL/6J males, while $+/+$ embryos were obtained from matings of outbred albino CD-1 (Charles Rivers) males and females. Pseudo-pregnant females were obtained from matings of CD-1 females with vasectomized and proven sterile CD-1 males. All matings were spontaneous. Female mice were killed on the day of vaginal plug detection (day 1 of pregnancy) and their oviducts excised. The ampulla of the oviduct was punctured with watchmaker forceps and the released embryos were freed from surrounding cumulus cells by incubation in modified Whitten's medium^{10,11} containing 500 U ml⁻¹ bovine hyaluronidase (Sigma). Nuclear transplantation was performed as previously described⁵. Among 545 embryos, 524 (96%) were successfully enucleated and of the 524 pronuclear karyoplasts, 494 (94%) were successfully injected along with inactivated Sendai virus into the perivitelline space of enucleated recipient embryos. 403 (82%) of the pronuclear karyoplast:enucleated cytoplasm pairs underwent fusion. The latter were transferred on the day of microsurgery to the oviducts of day 1 pseudo-pregnant females using a small-bore glass pipette.

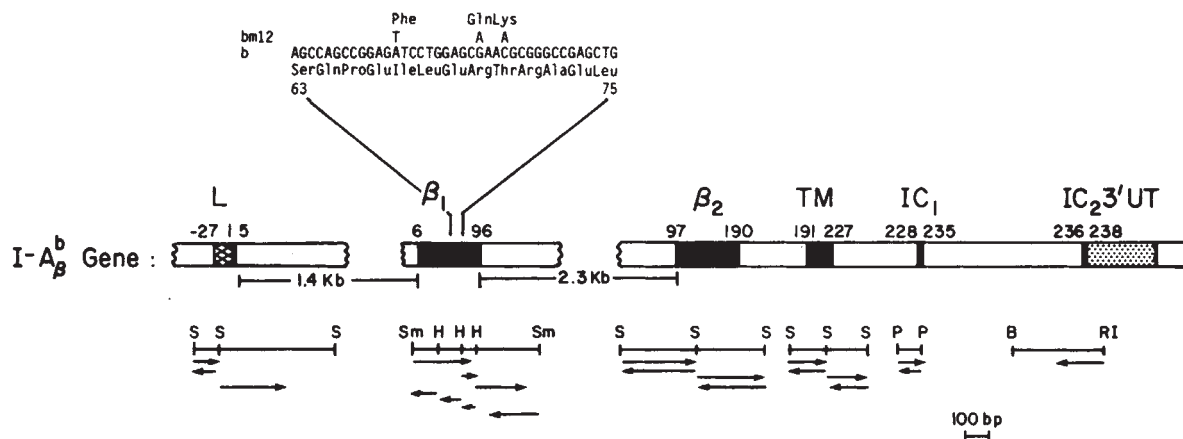


Fig. 1 The structure and sequencing strategy of the *I-A^β*^b and mutant *I-A^β*^{bm12} genes of the mouse, with coding region differences indicated. The exons of the *I-A^β*^b gene are denoted by solid shading in the regions encoding the mature polypeptide, by hatched shading in the region encoding the signal peptide, and by stippled shading in the 3' untranslated (3'UT) region. The exons are labelled as follows: L encodes the signal peptide and first five amino acids of the mature polypeptide, β_1 and β_2 encode the two extracellular domains, TM encodes the transmembrane region, and IC_1 and IC_2 encode the intracytoplasmic region. The 3'UT region is contiguous with IC_2 . The numbers above the shaded regions indicate the amino acid residues encoded by that exon. From bm12 genomic DNA, the 6.5-kb *EcoRI* fragment that carries most of the *I-A^β*^{bm12} gene was cloned into bacteriophage λ and subsequently into the plasmid vector pBR327 and identified by hybridization to a 2.1-kb *BamHI* fragment derived from the *I-A^β*^b gene. (The procedures for cloning a single-copy restriction fragment have been described previously²⁶.) The pBR327 clone of *I-A^β*^{bm12} and a pBR322 clone⁴ of *I-A^β*^b were digested with the restriction enzyme *SmaI*, *HpaII*, *Sau3AI*, *PstI*, *BamHI* or *EcoRI* and subcloned into M13 mp9 for nucleotide sequencing by the dideoxy chain termination method^{27,28}. The M13 recombinants were screened by a combination of hybridization to a human HLA-DC β -chain cDNA clone²⁹ and dideoxy chain termination 'single-lane screening' to identify among the M13 clones derived from the mutant *I-A^β*^{bm12} gene those which corresponded to previously isolated⁴ M13 clones derived from the wild-type *I-A^β*^b gene. M13 clones carrying analogous inserts derived from the wild-type and mutant genes were analysed together; that is, the products of the sequencing reactions for the two clones were electrophoresed in parallel to allow direct comparison of the two sequence ladders on the same gel. The sequencing strategy for the wild-type and mutant genes is presented below the structure of the gene. The restriction sites are designated as follows: Sm, *SmaI*; H, *HpaII*; S, *Sau3AI*; P, *PstI*; B, *BamHI*; RI, *EcoRI*. Arrows indicate the extent of reading from a given restriction site. The nucleotide stretch encoding amino acids 63-75 has been enlarged above the structure of the gene. Indicated above the b haplotype sequence are the three coding positions at which the bm12 haplotype sequence differs. Above each of these three nucleotide substitutions is indicated the amino acid substitution which it engenders in the *I-A^β* polypeptide. Sequencing of the six exons indicated in the gene structure demonstrated no additional nucleotide differences between the wild-type and mutant genes in their coding regions.

The coding regions of the *I-A^β*^b and *I-A^β*^{bm12} genes were subjected to nucleotide sequence analysis (Fig. 1). No differences were found between the two genes in any of their coding blocks except the β_1 exon, where three nucleotide substitutions were observed within a stretch of 14 nucleotides (Figs 1, 2). All three substitutions have been confirmed by sequencing of both DNA strands in the region containing the mutation. Two out of the three nucleotide replacements affect restriction sites: A $\rightarrow$ T leads to the loss of a *Sau3AI* (TGAATC) site, while C $\rightarrow$ A leads to the loss of a *ThaI* (CGVCG) site. Both of these predicted changes in restriction sites have been confirmed by Southern blot analysis of the plasmid subclones carrying the *I-A^β*^b and *I-A^β*^{bm12} genes (data not shown).

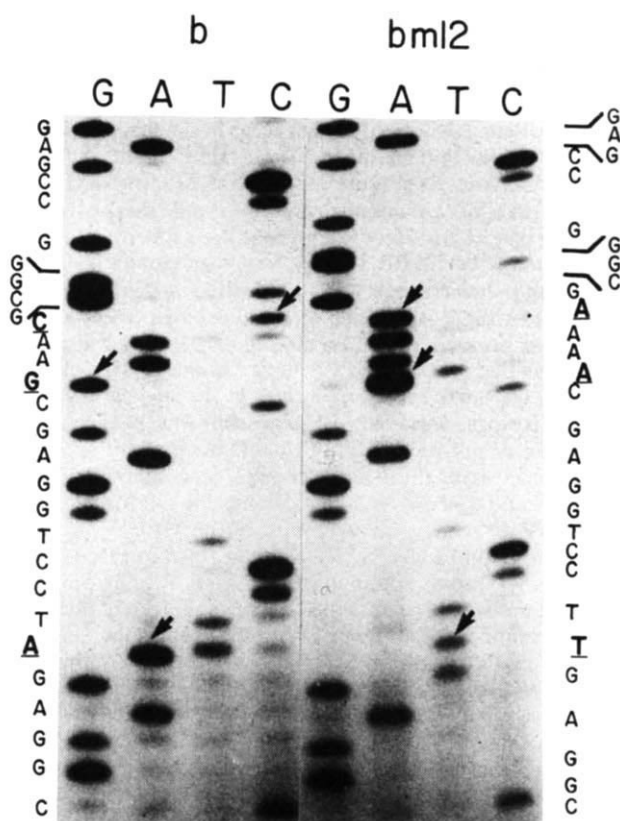


Fig. 2 Sequencing gel demonstrating nucleotide differences between *I-A^β*^b and mutant *I-A^β*^{bm12} genes. A pair of M13 clones, each carrying a *HpaII* insert derived from the β_1 exon of the *I-A^β*^b or *I-A^β*^{bm12} gene, was sequenced by the dideoxy chain termination method^{27,28}. The products of the sequencing reactions were electrophoresed in parallel on an 8% polyacrylamide sequencing gel. Reaction lanes (G, A, T, C) are designated at the top. The nucleotide sequences read off this gel are indicated at the sides, with the direction of reading from bottom to top corresponding to the 5' to 3' direction along the coding strand, starting with the first nucleotide of the insert. Nucleotide differences between the two sequences are underlined. Arrows mark the unique gel bands corresponding to these nucleotide differences. All three differences have been confirmed by sequencing of the pair of M13 clones carrying this same *HpaII* insert in the reverse orientation (data not shown). (The appearance of 'ghost' bands in the T lane at position 6 of the b ladder and at positions 5 and 16 of the bm12 ladder, as well as the 'compression' of bands seen at positions 22 and 23 of the b ladder and at position 23 of the bm12 ladder, are sequencing artefacts which are controlled for by sequencing of the complementary strand.)

I-A(b) (6)	HFVYQFMGECYFTNGTQRIRYVTRYIYNREEYVRYSVDVGEHRAVTELGRPDAEYWNQSPEILERTRAELDTVCRHNYEGPETHTSLRRLE	(96)
I-A(bm12)		
I-A(d)	V K Y L Y F QK V A S	
I-A(k)	H QPF L I F Y K -Y- KT P	
I-E(d)	R LEYVTS H Y HV FLE F NL F Y N DA SV Y ISDKFLVR V-	
DC	D K M EV LSS S V F F L L A KD Q AV R QLELRT LQ V	
DR2,2	R LW PKR H F E V FLD F Q S F F Q AV Y GVV SF VQ V-	
DR4,w6	R LELLKS H F E V FLE HFH Q A F Y R KDL QK GQV NY GVV SF VQ V-	
SB	***** E KD EE VP RM LGGPM LQ V-	

Fig. 3 Amino acid sequence comparison of the β_1 domains of mouse (I-A and I-E) and human (HLA-DC, -DR, and -SB) class II antigen β chains. The protein sequence of the β_1 domain (amino acids 6–96) of the I-A β^b chain, as deduced from the nucleotide sequence⁴, is shown in the one-letter amino acid code. Amino acid residues of the β_1 domain that differ from the I-A β^b sequence are shown below for the other eight β chains. Dashes represent gaps inserted into a sequence to maintain maximal homology. An asterisk indicates that the amino acid at that position has not been determined. For two out of the three positions at which the bm12 haplotype I-A β sequence differs from the *b* haplotype sequence, a human β -chain sequence is known which has one or both of the same amino acids (underlined) as are found in the bm12 sequence. The β -chain sequences for this comparison were taken from the following sources: I-A β^d and I-A β^k from Choi *et al.*⁴, I-E β^d from Saito *et al.*¹⁸, DC from Larhammar *et al.*²⁹, DR2,2 from Kratzin *et al.*³⁰, DR4,w6 from Long *et al.*³¹ and SB from Roux-Dosseto *et al.*³².

Comparison of the nucleotide sequences of the I-A β^b and mutant I-A β^{bm12} genes indicates that complete tryptic cleavage of their protein products would generate identical peptides except for two peptides unique to I-A β^b which would be replaced by one peptide (plus an additional residue of free arginine) unique to I-A β^{bm12} . The slight discrepancy with the results of comparative peptide mapping studies^{13,14} can be explained by incomplete tryptic cleavage (for example, trypsin sometimes fails to cleave between adjacent lysine and arginine residues, D. J. McKean, personal communication). The nonconservative character of two out of the three amino acid replacements found in the I-A β^{bm12} polypeptide (Arg → Gln and Thr → Lys; Fig. 1) probably results in significant changes in the folding of the I-A β polypeptide, its ability to associate with the I-A α chain, and the tertiary structure of the I-A molecule as a whole. This supposition is in keeping with the well-documented antigenic disparity between I-A β^b and I-A β^{bm12} molecules^{5,6,11,12}. Furthermore, the finding that bm12 splenocytes express only 25–50% as much cell-surface I-A as do B6 splenocytes^{6,11,13} may reflect the reduced ability of the mutant bm12 I-A β chain to associate with the I-A α chain.

Clustered mutations that occur in a single event are generally thought to result from a gene conversion event or from the reciprocal exchange of genetic information between closely related sequences^{16,17}. To establish that the bm12 mutation (a spontaneous mutation) resulted from a gene conversion-like event, one must establish that a likely donor sequence exists in the mouse genome. At present, such a donor sequence has not been identified. However, the observation that two out of the three amino acid substitutions found in I-A β^{bm12} relative to I-A β^b (glutamine at position 70 and lysine at position 71) are also found at the same positions in some human class II antigen β chains (Fig. 3) suggests the possibility that another class II β -chain gene acted as a donor sequence for the bm12 mutation. The original bm12 mutant, a (B6 × BALB/c)F₁ mouse⁵, probably resulted from the fusion of an abnormal gamete from the B6 parent with a normal gamete from the BALB/c parent, and therefore a candidate donor sequence for gene conversion must be sought in the B6 genome. As only two class II β chains, I-A and I-E, have been identified in the mouse, the obvious candidate for a donor sequence for the bm12 mutation is the I-E β^b gene. However, the only I-E β sequence available for comparison at present is from the d haplotype (BALB/c)¹⁸. While the I-E β^d and I-A β^{bm12} sequences differ at all three positions affected by the bm12 mutation (Fig. 3), the overall nucleic acid homology between I-E β^d and I-A β^b in the β_1 exon is 78%, a value in keeping with our hypothesis, as sequence homology is thought to be a prerequisite for gene conversion^{16,17}.

The bm12 strain had its origins in an extensive screening programme for spontaneous H-2 mutants¹⁹, most of which were found to have an altered H-2K^b gene²⁰. Characterization of some of these H-2K^b mutations by amino acid and nucleotide sequence analyses suggested that they arose from gene conversion-like events involving donor sequences present in other class I genes^{21–24}. The results of the screening programme for H-2

mutants suggest that mutations in class I genes occur at a frequency of roughly 1 in 4,000 mice²⁰. By contrast, only a single mutation in a class II gene was identified on screening almost 100,000 mice. Thus, the frequency of mutation in class I and class II genes appears to differ by an order of magnitude, yet the mutational mechanism may be the same and the degree of polymorphism of the class I and II antigens is roughly the same²⁵. These results taken together suggest that neither gene conversion as a mutational mechanism nor the apparent mutation frequencies for class I and II genes can account for the extreme polymorphism of these genes. Instead, we suggest that selection is the major force governing the evolution and maintenance of this polymorphism.

We thank C. S. David and D. J. McKean for the bm12 mouse used in this study, P. A. Peterson for the HLA-DC β -chain cDNA clone²⁹ used to screen M13 recombinants, Karen Klein for technical assistance, Terri Broderick for editorial assistance, and Ed Choi and David McKean for helpful discussions. K.R.M. was supported by a postdoctoral fellowship (PF-02191) from the American Cancer Society. This research was supported by NIH grant AI-18436 and American Cancer Society grant JFRA-66 to J.G.S.

Received 24 October 1983; accepted 19 January 1984.

- Uhr, J. W., Capra, J. D., Vitetta, E. S. & Cook, R. G. *Science* **206**, 292–297 (1979).
- Benacerraf, B. *Science* **212**, 1229–1238 (1981).
- Benoist, C. O., Mathis, D. J., Kanter, M. R., Williams, V. E. II & McDevitt, H. O. *Cell* **34**, 169–177 (1983).
- Choi, E., McIntyre, K., Germain, R. N. & Seidman, J. G. *Science* **221**, 283–286 (1983).
- McKenzie, I. F. C. *et al. J. exp. Med.* **150**, 1323–1338 (1979).
- McKenzie, I. F. C., Sandrin, M. S., Morgan, G. M., Henning, M. M. & Melvold, R. W. *Immunogenetics* **11**, 103–106 (1980).
- Michaelides, M. *et al. J. exp. Med.* **153**, 464–469 (1981).
- Lin, C.-C. S., Rosenthal, A. S., Passmore, H. C. & Hansen, T. H. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6406–6410 (1981).
- Krcso, C. J., Kazim, A. L., Atassi, M. Z., Melvold, R. W. & David, C. S. *J. Immunogenet.* **8**, 471–476 (1981).
- Lei, H.-Y., Melvold, R. W., Miller, S. D. & Waltenbaugh, C. J. *exp. Med.* **156**, 596–609 (1982).
- Lafuse, W. P., McCormick, J. F., Melvold, R. W. & David, C. S. *Transplantation* **31**, 434–438 (1981).
- Hansen, T. H., Walsh, W. D., Ozato, K., Arn, J. S. & Sachs, D. H. *J. Immun.* **127**, 2228–2231 (1981).
- McKean, D. J., Melvold, R. W. & David, C. *Immunogenetics* **14**, 41–51 (1981).
- Lee, D. R., Hansen, T. H. & Cullen, S. E. *J. Immun.* **129**, 245–251 (1982).
- Larhammar, D. *et al. Cell* **34**, 179–188 (1983).
- Baltimore, D. *Cell* **24**, 592–594 (1981).
- Egel, R. *Nature* **290**, 191–192 (1981).
- Saito, H., Maki, R. A., Clayton, L. K. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5520–5524 (1983).
- Kohn, H. I. *et al. Immunogenetics* **7**, 279–294 (1978).
- Nairn, R., Yamaga, K. & Nathenson, S. G. *A. Rev. Genet.* **14**, 241–277 (1980).
- Evans, G. A., Margulies, D. H., Camerini-Otero, R. D., Ozato, K. & Seidman, J. G. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1994–1998 (1982).
- Pease, L. R., Schulze, D. H., Pfaffenbach, G. M. & Nathenson, S. G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 242–246 (1983).
- Weiss, E. H. *et al. Nature* **301**, 671–674 (1983).
- Mellor, A. L., Weiss, E. H., Ramachandran, K. & Flavell, R. A. *Nature* **306**, 792–795 (1983).
- Klein, J. *Science* **203**, 516–521 (1979).
- Seidman, J. G. & Leder, P. *Nature* **276**, 790–794 (1978).
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. *J. molec. Biol.* **143**, 161–178 (1980).
- Messing, J., Crea, R. & Seeburg, P. H. *Nucleic Acids Res.* **9**, 309–321 (1981).
- Larhammar, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 3687–3691 (1982).
- Kratzin, H. *et al. Hoppe Seyler's Z. physiol. Chem.* **362**, 1665–1669 (1981).
- Long, E. O., Wake, C. T., Gorski, J. & Mach, B. *EMBO J.* **2**, 389–394 (1983).
- Roux-Dosseto, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 6036–6040 (1983).

Recent gene conversion involving bovine vasopressin and oxytocin precursor genes suggested by nucleotide sequence

Siegfried Ruppert, Gerd Scherer & Günther Schütz

Institute of Cell and Tumor Biology, German Cancer Research Center, Im Neuenheimer Feld 280, D-6900 Heidelberg, FRG

The nonapeptide hormones arginine vasopressin (AVP) and oxytocin (OT) are synthesized in the hypothalamus together with their carrier proteins, the neurophysins, as common polypeptide precursors¹⁻⁴. The organization of these precursors has been established by sequence determination of cloned bovine cDNAs encoding prepro-arginine vasopressin-neurophysin II (prepro-AVP-NPII) and prepro-oxytocin-neurophysin I (prepro-OT-NPI)^{5,6}. When the mRNA sequences coding for the conserved middle part of the neurophysins were compared, we found that these sequences are not merely similar but identical⁶. The primary structure of the chromosomal genes now determined shows that both genes, which appear to have arisen by a gene duplication, are split into three exons, each encoding a functional domain of the precursor polypeptide. Sequence comparison reveals that the stretch of sequence identity within the two mRNAs is probably the result of a gene conversion encompassing exon B, which encodes the conserved part of the neurophysins, and part of the preceding intron.

A bacteriophage library of bovine genomic DNA was constructed and screened for AVP and OT sequences (Fig. 1). Five independent phage clones were isolated (three vasopressin

specific and two oxytocin specific). As the established restriction maps of these clones do not overlap, no conclusion can be derived as to whether or not these genes are closely linked in the bovine genome. The primary structure of the prepro-AVP-NPII and prepro-OT-NPI genes was determined from sequence analysis of appropriate restriction fragments. The consistency of the chromosomal gene sequences (Figs 1, 2) with the previously established mRNA sequences^{5,6} suggests that we have cloned the functional genes. This is supported by genomic Southern blots which gave no evidence for additional neurophysin-related sequences (not shown).

The overall structures of the bovine prepro-AVP-NPII and prepro-OT-NPI genes are very similar (Fig. 3). They both consist of three exons and two introns which occur at identical positions with respect to the reading frame (between codons 21 and 22 and within codon 89). As a reflection of their common history, in both genes each functional domain of the precursor peptide is encoded by separate exons. The signal peptide together with the hormones AVP and OT are encoded by the first exon (the hormone domain). The conserved middle part of the neurophysins I and II (amino acids 22-86 in Fig. 3, corresponding to residues 10-74 in the mature neurophysins) is located in the second exon (the neurophysin domain). In the AVP-NPII precursor gene the third exon carries the information for the glycoprotein (the glycoprotein domain). The variable amino-terminal part of the neurophysins (the first 9 amino acids) are within the first exon, whereas the hypervariable carboxy-terminal part is encoded jointly by the second (amino acids 87 and 88) and third exons (residues 89-107 for neurophysin II and 89-106 for neurophysin I). This similarity in gene structure is also seen in the corresponding rat genes^{7,30}.

When the 5'-flanking sequences of the bovine genes are compared with their rat counterparts, extended regions can be seen

Fig. 1 Primary structure of the bovine prepro-OT-NPI gene. The nucleotide sequence of the prepro-OT-NPI gene was determined by sequence analysis of appropriate restriction fragments derived from the oxytocin-specific phage clone λ bOT-NPI. This clone carries a 13.5 kb insert of bovine genomic DNA consisting of 5.0 kb of 5' flanking DNA, a 1.0 kb long structural gene region and 7.5 kb of 3' flanking DNA. Nucleotide residues are numbered below the sequence, with negative numbers assigned to the residues upstream of the +1 transcription initiation site indicated by the arrow. The modified Goldberg-Hogness box and the polyadenylation sequence are underlined. The poly(A) addition site is indicated by an arrow. Oxytocin is encoded by amino acid residues +1 to +9, neurophysin I (NPI) by amino acid residues +13 to +106. The putative signal peptide starts at amino acid -19. The box shows the region of perfect sequence homology to the AVP-NPII precursor gene (see Fig. 2). The sequence shown differs from the cDNA sequence reported previously⁶ only in the 3' noncoding region (position +869, C instead of T).

Methods: To construct a bovine genomic library high molecular weight DNA (80-100 kb) from fresh calf thymus was partially digested with *Sau*3A and fragments 12-20 kb in length were enriched by NaCl gradient centrifugation²⁶. Isolated fragments were ligated into the *Bam*HI sites of the bacteriophage vector EMBL3²⁷. Using packaging extracts (provided by V. Pirrotta) a total of 12×10^6 recombinant bacteriophages were generated. More than 98% of bacteriophages are recombinants and contain bovine DNA fragments of an average size of 14 kb. 5×10^6 original recombinants were screened for AVP and OT sequences according to the method of Benton and Davis²⁸, and subsequently amplified to establish a library. Three independent vasopressin specific clones were isolated containing the same *Sau*3A partial genomic fragment, in either orientation relative to the phage vector arms. The two oxytocin specific clones isolated also contain an identical *Sau*3A partial fragment, cloned in either orientation. The following probes were used. Plasmid pVNPII-1⁵ contains most of the mRNA sequences of prepro AVP-NPII; this probe also hybridizes to oxytocin specific clones because of the identical 197 bp sequence within the coding regions of NPI and NPII⁶. Plasmid p3'OTNPI contains the 3' *Pst*I fragment of pONPI-1⁶ and does not hybridize to prepro AVP-NPII cDNA sequences. DNA sequencing was carried out according to the procedure of Maxam and Gilbert²⁹. For most of the sequence both strands were sequenced. The start sites of transcription were determined by *S*₁ nuclease mapping experiments.

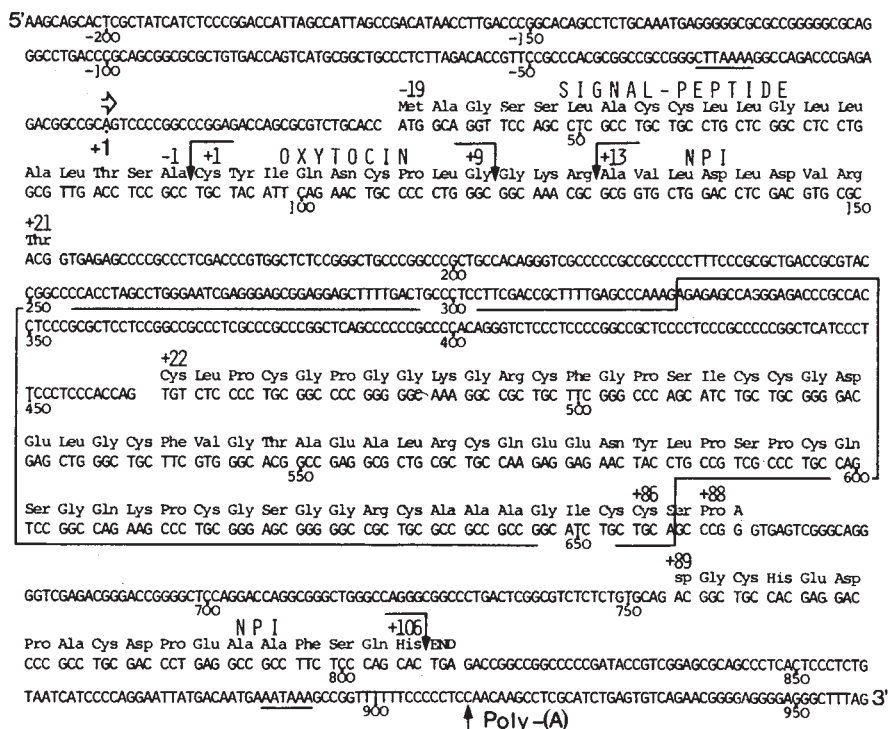
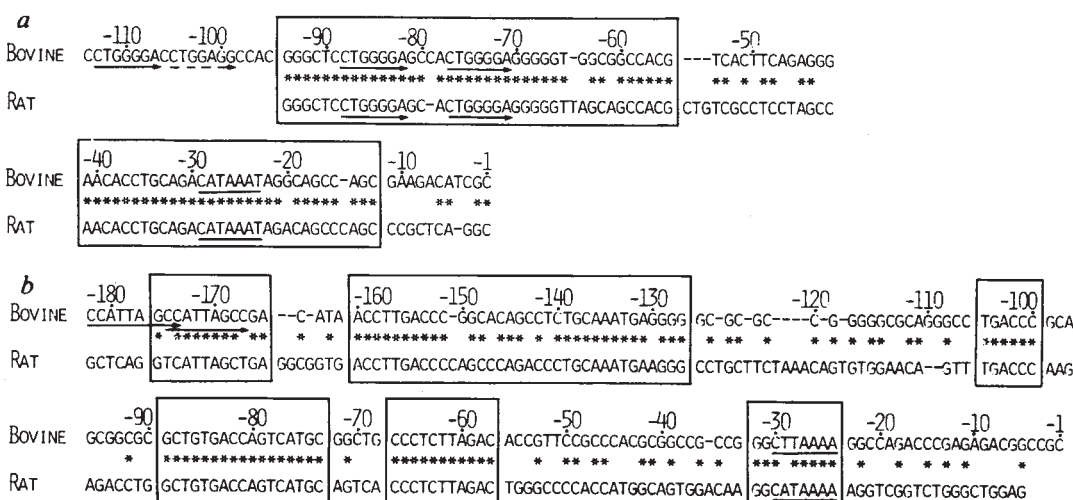


Fig. 4 Sequence comparison of the promoter region of the bovine and rat prepro-AVP-NP_{II} and -OT-NP_I genes. **a**, The sequences of the bovine and rat⁷ AVP-NP_{II} 5'-flanking regions are aligned to give the best match between the sequences. Two closely matching regions ranging from -12 to -41 and from -55 to at least -94 are indicated by boxed areas (no further data are available for the rat sequence). The sequences CTGGGGA found as direct repeats are indicated by arrows. A variant of this sequence, CTGGAGG, is indicated by a dashed arrow. **b**, Alignment of the 5' flanking bovine and rat³⁰ OT-NP_I sequences. Homologous regions ranging from -24 to -32, -57 to -67, -73 to -89, -100 to -105, -127 to -160 and -165 to -176 are indicated by boxes. The direct repeats CCATTAGCC are indicated by arrows. The numbering in **a** and **b** refers to positions within the bovine genes starting upstream of the initiation site. The modified Goldberg-Hogness boxes are underlined. Homologous nucleotides are indicated by asterisks.



where the sequences are very similar. Within the AVP-NP_{II} 5'-flanking regions two very well conserved blocks are found. Both within and 5' to the second block the heptanucleotide CTGGGGA, and a variant, CTGGAGG, occurs four times in the bovine gene and twice in the rat gene (Fig. 4a). Alignment of the OT-NP_I 5'-flanking regions reveals six conserved blocks (Fig. 4b). The sequence CCATTAGCCATTAGCC, a direct overlapping repeat, is found 5' to and within the sixth block of the bovine gene. These gene-specific homologies conceivably reflect the expression of these two genes by separate populations of neurones¹.

We previously observed that the bovine mRNA sequences coding for the conserved part of the neurophysins are perfectly identical over a stretch of 197 base pairs (bp)⁶. We suggest that the most likely explanation for this sequence identity within this gene pair, which arose by gene duplication about 450 Myr ago⁸, is a recent gene conversion^{9,10}. Such conversion events have recently been described in a number of genes, notably the globin¹¹⁻¹⁴, immunoglobulin^{9,15-17} and major histocompatibility antigen genes^{18,19}. We note that the 197 bp perfect sequence identity is located in the second exon and extends without interruption for 135 bp into the preceding intron. The perfect sequence identity between the two genes covers a total of 332 bp and breaks off close to the exon B/intron 2 junction (Figs 1, 2). The 60 bp preceding this converted region show 81% sequence identity while the next 100 bp upstream can only poorly be aligned (40% identity). It is possible that the converted segment originally included these additional 60 bp which experienced more mutational drift than the remaining 332 bp. Alternatively, the gene conversion event leading to the 332 bp perfect homology may have occurred very recently in evolution within an earlier gene conversion unit encompassing these additional 60 bp, suggesting that this region might represent a hot spot for gene conversion within the AVP/OT gene pair. Other than a high G+C content (77%) no obvious sequence abnormalities have been found which may have facilitated a gene conversion event, as has been observed in other cases^{11,12,14}.

Nucleotide substitutions leading to amino acid replacement provide a good evolutionary clock with a particular rate for each specific protein. In contrast, silent substitutions accumulate roughly independently of the rate at which the protein evolves^{18,20-22}. Percentage divergence of replacement (R) and silent (S) substitutions was calculated according to Lomedico *et al.*²⁰ for the bovine AVP/OT gene pair. Different values for replacement changes in the parts of exons A and C that encode for

different functional parts of the protein precursor are obtained while the accumulation of silent changes is similar (R: 24.4, 9.3, 45.5, 44.9; S: 56, 54.5, 41.7, 57.4 for signal peptide plus Gly-Lys-Arg linker, hormone, N-terminal part of neurophysin, and C-terminal part of neurophysin, respectively). Naturally, the converted exons B have very low silent (2.1) and replacement (1.9) changes.

In comparing the AVP-NP_{II} and OT-NP_I exons of the rat^{7,30} we again notice different mutation rates at replacement and similar rates at silent sites within the various functional parts of exons A and C (R: 29.8, 14, 38.5, 32.7; S: 53.9, 54.5, 64.5, 57.4). In exon B replacement changes are low as expected (2.6), but silent changes are also low (12.6). This 4-5-fold slower rate of silent substitutions in exon B (72% G/C) compared to exons A (61% G/C) and C (68% G/C) cannot completely be explained by a lower substitution rate due to a higher G/C content as observed by Smithies *et al.*²³ and suggests that an analogous but more remote gene conversion event involving exon B might also have occurred in the rat. On the basis of the mutation rate for silent substitutions calculated by Perler *et al.*²¹, we can estimate that the gene conversion in the rat genes occurred approximately 10 Myr ago and in the bovine genes less than 1 Myr ago.

Amino acid sequence comparisons of the neurophysins in various species²⁴ reveal strong conservation of that part of the neurophysin encoded in exon B, suggesting the functional importance of this part of the protein. Intergenic conversion of members of a gene family may guarantee coevolution of a sequence in closely related genes and thus may strongly contribute to the conservation of vital function²⁵. Alternatively, the conservation in amino acid sequence may simply be the accidental consequence of repeated gene conversions.

We thank V. Pirrotta for gifts of packaging extracts, S. Suhai for help with computer assisted sequence analysis, R. Ivell and D. Richter for making their rat prepro-OT-NP_I genomic DNA sequence available before publication, our colleagues for helpful comments on the manuscript, in particular W. Keller, H. Lehrach, R. Miksicek, R. Renkawitz and A. Sippel, and M. Cole for typing. This work was supported by Fonds der Chemischen Industrie.

Received 21 October 1983; accepted 9 January 1984.

- Zimmermann, E. A. & Defendini, R. in *Neurohypophysis* (eds Moses, A. M. & Share, L.) 22-29 (Karger, Basel, 1977).
- Brownstein, M. J., Russell, J. T. & Gainer, H. *Science* **207**, 373-378 (1980).
- Sachs, H., Fawcett, P., Takabatake, Y. & Portanova, R. *Recent Prog. Horm. Res.* **25**, 447-491 (1969).

4. Schmale, H. & Richter, D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 766–769 (1981).
5. Land, H., Schütz, G., Schmale, H. & Richter, D. *Nature* **295**, 299–303 (1982).
6. Land, H. *et al. Nature* **302**, 342–344 (1983).
7. Schmale, H., Heinssohn, S. & Richter, D. *EMBO J.* **2**, 763–767 (1983).
8. Acher, R. *Proc. R. Soc. B210*, 21–43 (1980).
9. Baltimore, D. *Cell* **24**, 592–594 (1981).
10. Petes, T. & Fink, G. R. *Nature* **300**, 216–217 (1982).
11. Slightom, J. L., Blechl, A. E. & Smithies, O. *Cell* **21**, 627–638 (1980).
12. Shen, S., Slightom, J. L. & Smithies, O. *Cell* **26**, 191–203 (1981).
13. Liebhaber, S. A., Goossens, M. & Kan, Y. W. *Nature* **290**, 26–29 (1981).
14. Schon, E. A., Wernke, S. M. & Lingrel, J. B. *J. biol. Chem.* **257**, 6825–6835 (1982).
15. Bentley, D. L. & Rabbitts, T. H. *Cell* **32**, 181–189 (1983).
16. Ollo, R. & Rougeon, F. *Cell* **32**, 515–523 (1983).
17. Schreier, P. H., Bothwell, A. L. M., Müller-Hill, B. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4495–4499 (1981).
18. Hayashida, H. & Miyata, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2671–2675 (1983).
19. Weiss, E. H. *et al. Nature* **301**, 671–674 (1983).
20. Lomedico, P. *et al. Cell* **18**, 545–558 (1979).
21. Perler, F. *et al. Cell* **20**, 555–566 (1980).
22. Elstratidis, A. *et al. Cell* **21**, 653–668 (1980).
23. Smithies, O., Engels, W. R., Devereux, J. R., Slightom, J. L. & Shen, S. *Cell* **26**, 345–353 (1981).
24. Chauvet, M. T., Hupert, D., Chauvet, J. & Acher, R. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2839–2843 (1983).
25. Hood, L., Campbell, J. H. & Elgin, S. C. R. *A. Rev. Genet.* **9**, 305–353 (1975).
26. Liu, C., Tucker, P. W., Mushinski, F. & Blattner, F. R. *Science* **209**, 1348–1353 (1980).
27. Frischauf, A.-M., Lehrach, H., Poustka, A. & Murray, N. *J. molec. Biol.* **170**, 827–842 (1983).
28. Benton, W. D. & Davis, R. W. *Science* **196**, 180–182 (1977).
29. Maxam, A. M. & Gilbert, W. in *Methods Enzym.* **65**, 499–560 (1980).
30. Ivell, R. & Richter, D. *Proc. natn. Acad. Sci. U.S.A.* (in the press).

Direct measurement of picosecond charge separation in bacteriorhodopsin

G. I. Groma*, G. Szabó† & Gy. Váró*

* Institute of Biophysics, Biological Research Centre of Hungarian Academy of Sciences, Odessai krt. 62, Szeged 6726, Hungary

† Institute of Experimental Physics, József Attila University of Sciences, Dóm tér 9, Szeged 6720, Hungary

The *Halobacterium halobium* protein bacteriorhodopsin conserves the energy of absorbed photons by converting it into a transmembrane proton gradient¹. Light absorption by bacteriorhodopsin is thought to drive a photocycle of intermediate states linked to the pumping of protons across the plasma membrane. The earliest intermediate of this photocycle so far detected is formed in 11–15 ps^{2,3}, and this step could involve a separation of charges within the protein^{2,4–9}. Although an electrical response signal with a time course correlating with that of the photocycle has been measured for bacteriorhodopsin, so far it has not been possible to resolve a signal corresponding to the initial charge separation^{10–16}. We report here the resolution by picosecond laser spectroscopy of an electrical signal apparently corresponding to a charge separation with a time constant of approximately 30 ps, which we attribute to the formation of the first intermediate of the bacteriorhodopsin photocycle.

Experiments were carried out on dried oriented purple membrane samples as previously described¹⁷. The membrane layers were fixed on glass plates covered by an electrically conducting and optically transparent SnO₂ thin layer. The samples were mounted onto a 50-Ω coaxial connector. The outer electrode of this connector was connected to the SnO₂ layer, while the inner one was tightly attached to the top of the membrane layer.

Figure 1 shows the experimental setup. The samples were exposed by an N₂ laser pumped distributed feedback dye laser (DFDL)^{18,19}. This new, very versatile and inexpensive picosecond laser source provided single pulses of 55 ps and 50 μJ energy at 10 Hz repetition rate. The electric response signal of the sample was directly fed into the input of a 14-GHz sampling oscilloscope (Tektronix 7704A with S-4 sampling head, rise time 25 ps) via the 50-Ω connector. It was triggered by an ultrafast photodiode (Instrument Technology, UK, type HSD50, risetime 100 ps). The oscilloscope output was digitalised and stored in a multichannel analyser (ICA 70, KFKI, Hungary). Light adaptation of the sample was ensured by preillumination with the exciting laser. Experiments were at room temperature.

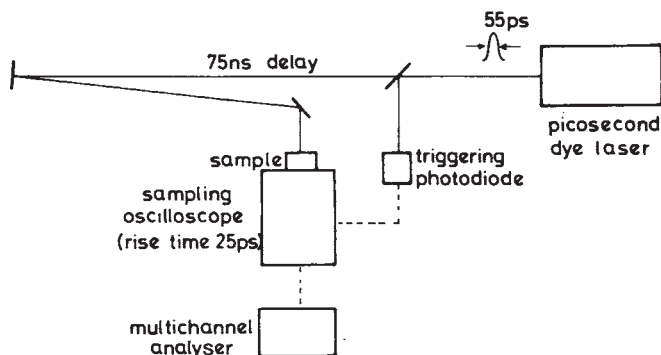


Fig. 1 Scheme of the experimental set-up.

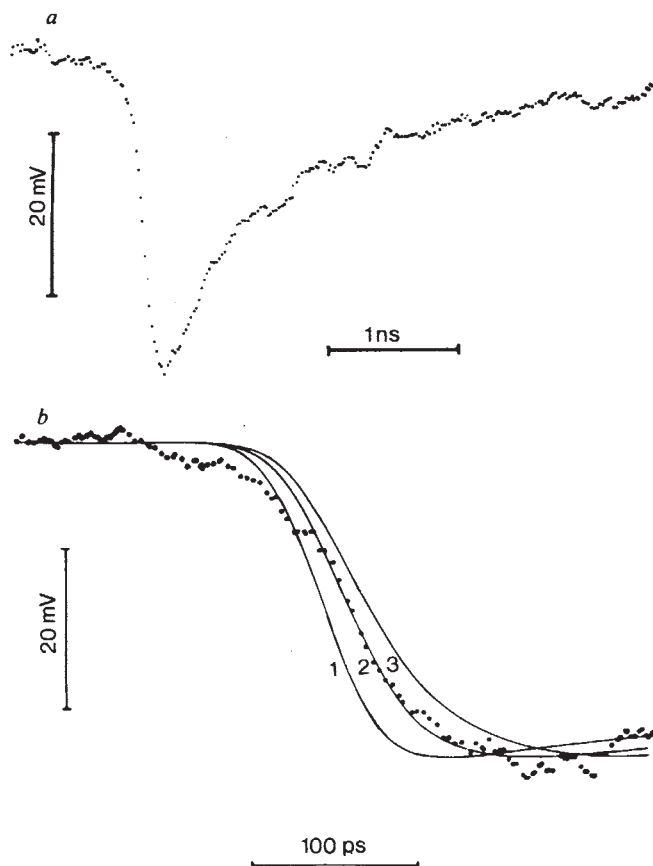


Fig. 2 a, The overall bacteriorhodopsin electric response signal. For experimental conditions see Fig. 1 and the text. Average of 20 scans. Dotted line: the rising part of the signal measured with higher time resolution. Average of 32 scans. Solid line: computer simulation of the signal using equation (1). $RC = 1,800$ ps (determined from the decay of the signal), (1) $\tau = 10$ ps, (2) $\tau = 30$ ps, (3) $\tau = 50$ ps.

A characteristic picosecond protein electric response signal (PERS) of bacteriorhodopsin is presented in Fig. 2. Like other workers^{10–16} we found that this signal is negative with respect to the overall proton pump, the apparent rise time of which is close to the laser pulse width and the oscilloscope rise time. For this reason we used network analysis to decompose PERS from the measured signal. Supposing a linear system the output signal $S(t)$ is determined by the convolution of the input and the weighting function of each subsystems²⁰:

$$S(t) = E(t) \otimes P(t) \otimes C(t) \otimes O(t) \quad (1)$$

where $E(t)$ is the time distribution of the exciting laser pulse described by a gaussian function with halfwidth of 55 ps, $P(t)$ is the PERS (current signal) described by $P(t) = \text{const exp}(-t/\tau)$, where τ is to be determined, $C(t)$ is the

weighting function of the measuring circuit described by $C(t) = \text{const exp}(-t/RC)$, where $R = 50 \text{ ohm}$ and C is $10\text{--}100 \text{ pF}$, $O(t)$ is the weighting function of the oscilloscope described by a gaussian function with halfwidth of 25 ps and $\otimes$ denotes the operation of convolution.

Note that equation (1) deals with time-independent bacteriorhodopsin concentration. We also made detailed calculations on more sophisticated nonlinear models including bacteriorhodopsin depletion and the back photoreaction of the first intermediate (K) to bacteriorhodopsin. However, they gave almost identical results at the laser intensity used, although big differences occurred at higher intensities.

Equation (1) cannot be directly solved by deconvolution methods because of the high noise level. Instead we used a computer simulation of $S(t)$ with a set of the probe value of τ and RC . The measured and computed curves are compared in Fig. 2b. Clearly the middle part of the measured signal fits well to the computed one with $\tau = 30 \text{ ps}$. The ringing at the ends is probably due to the non-ideal response of the oscilloscope (fast-rising signal in a band limited system) and the connector (for example, presence of inductivity). The step at the first third of the signal may represent a reflection but also a real very fast unresolved phase in the sample response. The decay of the curve on Fig. 2a is much slower than its rise. According to equation (1), that is mainly determined by the RC constant of the measuring circuit (for detailed analysis see ref. 10) and really varied with the capacitance of samples. No additional significant PERS could be observed on the decaying part and on any larger time scale with the low input impedance used.

Figure 2b suggests a charge separation process with a time constant of $30 \pm 10 \text{ ps}$. Considering the very close value of K formation time^{9,10} we attribute the measured PERS to this transition. Note, however, that the real frequency transfer limitation of the used sample and connector construction is unknown, so the observed charge separation actually can be even faster, correlating with any state preceding K formation.

The current models of bacteriorhodopsin to K transition^{2,4-9} suppose retinal isomerization and/or proton transfer. Both of them predict the existence of the observed signal. The data presented here are not enough to distinguish between them but ultrafast PERS seems to be a powerful tool for investigating the molecular mechanism of bacteriorhodopsin photochemistry. In addition, picosecond charge separation is not a unique feature of bacteriorhodopsin. Recently the observation of an electric signal faster than 200 ps was reported on chloroplasts²¹.

It is interesting to compare PERS of purple membrane with photocurrent generation in semiconductors. Both produce current in a measuring circuit upon illumination. Note, however, that the two systems work by crucially different mechanisms. The primary effect of light on semiconductors is only the generation of mobile charges. At the same time a primary charge separation in bacteriorhodopsin molecules occurs even on excitation by the polarization of retinal chromophore^{22,23} followed by isomerization or proton transfer. So light acts as a charge generator on semiconductors but as a direct displacement current generator on ordered bacteriorhodopsin molecules.

We thank Drs Gy. Farkas, Z. Horvath, I. Santa and J. Hebling for their help with the experiments and Professor L. Keszthelyi for fruitful discussions.

Received 8 August 1983; accepted 6 February 1984.

1. Stoeckenius, W., Lozier, R. & Bogomolni, R. *Biochim. biophys. Acta* **505**, 215-278 (1979).
2. Applebury, M., Peters, K. & Rentzepis, P. *Biophys. J.* **23**, 375-382 (1978).
3. Gillbro, T. & Sundström, V. *Photochem. Photobiol.* **37**, 445-455 (1983).
4. Lewis, A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 549-553 (1978).
5. Warshel, A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2558-2562 (1978).
6. Schulten, K. & Tavan, P. *Nature* **272**, 85-86 (1979).
7. Honig, B., Ebrey, T., Callender, R. & Ottolenghi, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2503-2507 (1979).
8. Gochev, A. & Christov, S. *Biophys. struct. Mech.* **7**, 187-193 (1981).
9. Birge, R. & Cooper, T. *Biophys. J.* **42**, 61-69 (1983).
10. Keszthelyi, L. & Ormos, P. *FEBS Lett.* **109**, 189-193 (1980).
11. Drachev, L., Kaulen, A. & Skulachev, V. *FEBS Lett.* **87**, 161-167 (1978).

12. Hong, F. & Montal, M. *Biophys. J.* **25**, 465-472 (1979).
13. Fahr, A., Läger, P. & Bamberg, E. *J. Membrane Biol.* **60**, 51-62 (1981).
14. Ormos, P., Reinisch, L. & Keszthelyi, L. *Biochim. biophys. Acta* **722**, 471-479 (1983).
15. Trissl, H.-W. *Biochim. biophys. Acta* **723**, 327-331 (1983).
16. Váró, Gy. & Keszthelyi, L. *Biophys. J.* **43**, 47-51 (1983).
17. Váró, Gy. *Acta Biol. Acad. Sci. Hung.* **32**, 301-310 (1981).
18. Bor, Zs. *IEEE J. Quantum Electron.* **QE-16**, 517-524 (1980).
19. Bor, Zs., Rácz, B., Szabó, G., Mueller, A. & Dorn, H. P. *Helv. phys. Acta* **56**, 383-392 (1983).
20. Fodor, Gy. *Laplace Transform in Engineering*, 43-46 (Hungarian Academy of Sciences, Budapest, 1965).
21. Trissl, H.-W., Kunze, U. & Junge, W. *Biochim. biophys. Acta* **682**, 364-377 (1982).
22. Salem, L. & Bruckmann, P. *Nature* **258**, 526-528 (1975).
23. Mathies, R. & Stryer, L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2169-2173 (1976).

Pumping of protons from the mitochondrial matrix by cytochrome oxidase

Mårten Wikström

Department of Medical Chemistry, University of Helsinki, Siltavuorenpenger 10A, SF-00170 Helsinki 17, Finland

The stoichiometry and mechanism of redox-linked proton translocation by the mitochondrial respiratory chain is a major issue of debate in membrane bioenergetics. The function of cytochrome oxidase is a focal point of disagreement. In 1977 it was suggested that the terminal component of the respiratory chain, cytochrome oxidase, functions as a redox-linked proton pump¹. That and subsequent studies were based mainly on measurements of proton ejection from mitochondria or from vesicles reconstituted with isolated cytochrome oxidase, or on measurements of translocation of electrical charge equivalents across mitochondrial and vesicle membranes²⁻⁶. This proton-translocating function of cytochrome oxidase is confirmed here by a quantitative determination of proton uptake from the inside (matrix) of intact mitochondria.

Several laboratories²⁻⁶ have reported evidence in support of the notion that cytochrome oxidase translocates protons across the mitochondrial inner membrane (see Fig. 1). However, the groups of Mitchell⁷ and Papa⁸⁻¹⁰ continue to maintain that cytochrome oxidase does not translocate protons. The groups of Azzone¹¹ and Lehninger¹², whilst supporting the proton-translocating activity ascribed to cytochrome oxidase, argue that the stoichiometry of proton-pumping is higher than that originally proposed and reported for vesicles containing cytochrome oxidase. On the other hand, there is full agreement that oxidation of succinate (or ubiquinol) by cytochrome *c* is associated with uptake of one proton from the matrix side per transferred electron ($1 \text{ H}^+/\text{e}^-$), as also depicted in Fig. 1.

Junge and his collaborators have shown that the pH-indicator Neutral red, which penetrates through membranes, responds to energization of the chloroplast^{13,14}. If the pH of the extravesicular solution is kept constant by a high concentration of an impenetrant buffer, the absorption changes of the dye reflect intravesicular changes in H^+ ion activity. In a similar way, Neutral red has also been used qualitatively to monitor pH changes in the matrix of isolated mitochondria^{1,15}. Although the indicator is mainly bound to the membrane of chloroplasts it nevertheless seems to report changes in pH of the intravesicular space, at least in the time range relevant for the present purposes^{13,16}. For mitochondria it is not yet proven that Neutral red records the intravesicular pH, but as discussed below the present work is independent of this assumption.

In the experiments reported here, the calibration of the Neutral red signal for matrix alkalization was carried out for oxidation of succinate by ferricyanide, taking advantage of the consensus that this reaction is linked to uptake of $1 \text{ H}^+/\text{e}^-$ (Fig. 1). Figure 2a shows absorption transients of Neutral red linked to reduction of calibrated small aliquots of ferricyanide by an

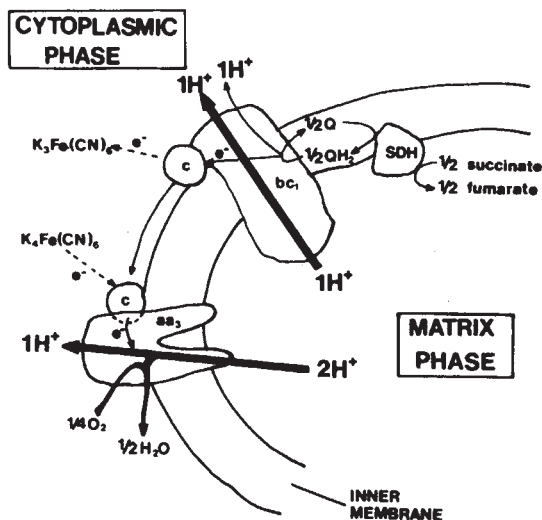


Fig. 1 Scheme of proton translocation by the cytochrome complexes of mitochondria. For the cytochrome bc_1 complex there is full agreement in the literature⁶. For cytochrome oxidase my original suggestion is shown^{1,2,6}. Oxidoreduction is described on a one-electron basis according to the electrochemical convention. Heavy arrows denote proton translocation. Electron flow through subsections of the electron transport system with ferricyanide as electron acceptor and ferrocyanide as electron donor is also shown. SDH, succinate dehydrogenase complex; Q, QH_2 , oxidized and reduced forms of ubiquinone; c, cytochrome c; aa_3 , cytochrome oxidase.

excess of succinate. Cyanide was present to block cytochrome oxidase. This experiment is directly comparable with the oxidant pulse method introduced by Mitchell *et al.*¹⁷, except that the intramitochondrial rather than the extramitochondrial pH is monitored. After consumption of the added ferricyanide within a few seconds, the trace relaxes to the baseline with kinetics similar to the relaxation of H^+ ejection (compare with ref. 17). As is the case for extramitochondrial H^+ ejection, the relaxation of the trace is greatly accelerated by proton-conducting uncoupling agents, which at high concentrations abolish net alkalization entirely (not shown). Figure 2b shows that the extent of absorption change is linearly dependent on the amount of added oxidant, up to at least 100 nmol ferricyanide in these conditions. As $1 H^+/e^-$ is known to be taken up from the matrix during oxidation of succinate by ferricyanide (Fig. 1), this yields a linear calibration of the absorption scale in terms of intramitochondrial ΔH^+ . The units of the abscissa of Fig. 2b can thus be replaced by nanomoles of H^+ ions.

The Neutral red absorption scale calibrated in terms of ΔH^+ was then used to determine quantitatively proton uptake from the matrix coupled to cytochrome c oxidase activity. Figure 3 (trace c) shows the kinetics of intramitochondrial alkalization during oxidation of ferrocyanide by O_2 , as catalysed by cytochrome c oxidase (see Fig. 1). Rotenone, antimycin and myxothiazol were present to block NADH dehydrogenase and the cytochrome bc_1 complex (see ref. 18). 80 nmol of ferricyanide was added first to exhaust endogenous reductants of ferricyanide, and to calibrate the absorption change at 420–460 nm (Fig. 3, trace a). Measurements of Neutral red absorption during these additions of ferricyanide (not shown) revealed no significant intramitochondrial alkalization. Thus, as expected, ferricyanide elicits no true proton translocation linked to the NADH–cytochrome c segment of the respiratory chain in the presence of rotenone, antimycin and myxothiazol. It is clear, therefore, that the proton translocation observed during oxidation of ferrocyanide (Fig. 3c) is due to cytochrome c oxidase, and not to reduction of ferricyanide by endogenous reductants as suggested by Moyle and Mitchell¹⁹.

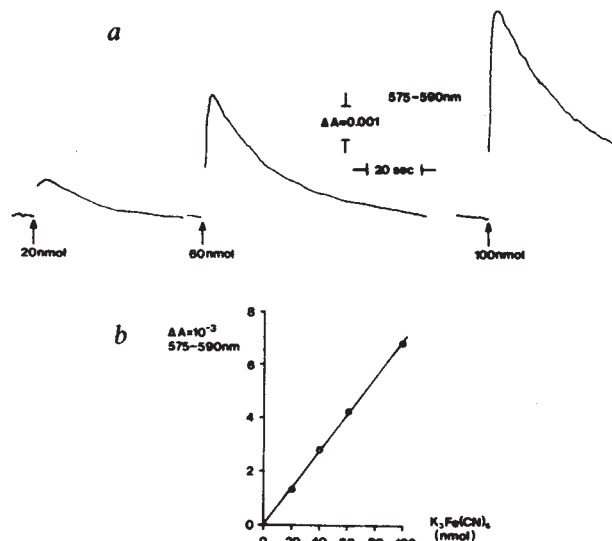


Fig. 2 Proton uptake from the mitochondrial matrix linked to reduction of ferricyanide by succinate. *a*, Traces obtained following additions of carefully calibrated aliquots of potassium ferricyanide. Upward deflections correspond to decreases in absorption at 575 nm relative to 590 nm. *b*, Absorption changes halfway in time between addition of the oxidant and its exhaustion (compare with ref. 17) relative to the amount of added ferricyanide.

Methods: The basic reaction medium contained 100 mM sucrose, 20 mM KCl, 50 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulphonate (HEPES buffer), adjusted to pH 7.4 with KOH. Further additions were 3 μ M rotenone, 210 μ M *N*-ethylmaleimide (to block backflux of H^+ in symport with inorganic phosphate; see refs. 1, 2, 6), 1 mM KCN, 0.5 μ g ml⁻¹ valinomycin (to prevent the formation of a membrane potential by making the membrane permeable to K^+), 83 μ M Neutral red and rat liver mitochondria equivalent to 0.83 μ M of cytochrome oxidase (~ 5.9 mg of protein ml⁻¹). After 2 min of incubation at 25 °C 4.2 mM potassium succinate was added, followed by further incubation for 5 min. The final volume was 1.2 ml.

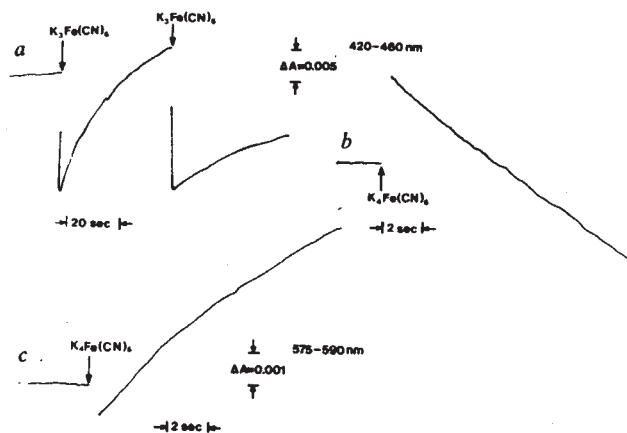


Fig. 3 Determination of electron transfer and mitochondrial matrix alkalization during ferrocyanide oxidation. *a*, Measurements of ferricyanide following two additions of 40 nmol aliquots of ferricyanide. *b*, Generation of ferricyanide following the initiation of respiration with 1.67 mM potassium ferrocyanide. *c*, Changes in Neutral red absorption following the initiation of respiration with 1.67 mM potassium ferrocyanide.

Methods: The conditions were the same as for Fig. 2, except that cyanide and succinate were omitted, and 0.25 μ g ml⁻¹ antimycin plus 4.6 μ M myxothiazol were added. Absorption changes of Neutral red were measured at 575 nm and 590 nm and ferricyanide at 420 nm and 460 nm. Upward deflections signify a decrease of absorption at the wavelength mentioned first, relative to that mentioned second.

Table 1 Proton uptake from the mitochondrial matrix

Condition	Respiratory rate (nmol e ⁻ s ⁻¹)	H ⁺ /e ⁻
As Fig. 3	5.4 (±0.2 s.d.)	-1.87 (±0.11 s.d.); n = 14
+4.2 mM MgCl ₂	16 (±0.9 s.d.)	-1.69 (±0.13 s.d.); n = 9

Experimental conditions were as described in the legend to Fig. 3. The data refer to experiments with eight different preparations of rat liver mitochondria. The H⁺/e⁻ ratios were obtained from initial rates of proton uptake (Fig. 3c) and electron transfer (Fig. 3b), which were calibrated for each preparation of mitochondria as shown in Figs 2 and 3a, respectively. Electron transfer between succinate and ferricyanide was assumed to result in uptake of 1 H⁺/e⁻ from the matrix phase (see Fig. 1 and the text).

Antimycin-insensitive reduction of ferricyanide by endogenous hydrogenated reductants can occur in some conditions^{19,20} (Fig. 3, trace a). This causes scalar release of H⁺ ions on the cytoplasmic side of the membrane, which should not be confused with true proton transport. During respiration with ferrocyanide (see Fig. 1) the rate of oxygen consumption (controlled separately with a Clark electrode; not shown, but see ref. 20) was found to match completely the net rate of generation of ferricyanide (Fig. 3, trace b), giving the electron transfer velocity. Without the pretreatment with ferricyanide there was in these conditions about 20% reduction in the net initial rate of production of ferricyanide relative to the consumption of O₂, due to reduction of formed ferricyanide by an endogenous reductant^{19,20} (see also Fig. 3, trace a).

When the rate of absorption change of Neutral red was converted to ΔH⁺ according to the calibration described above, it was found that the H⁺/e⁻ ratio of proton uptake was close to 2.0 during oxidation of ferrocyanide (Table 1). This confirms that generation of protonmotive force by the cytochrome oxidase reaction is twice as effective as that of succinate-cytochrome c reductase per transferred electron, in full agreement with my original proposal¹ (Fig. 1). A threefold increase in the rate of oxidation of ferrocyanide brought about by addition of MgCl₂ had little effect on the measured H⁺/e⁻ ratio (Table 1). Note that proton uptake and electron transfer were monitored from about 0.4 s onwards after addition of ferrocyanide (Fig. 3, traces b, c), that is, almost from the onset of e⁻ transport through cytochrome oxidase.

The possibility that changes in light scattering might significantly affect the measurements of Neutral red absorption was tested by using several different pairs of wavelengths, but this did not change the results. Selected pairs of wavelengths that gave absorption changes due to Neutral red and light scattering in the same or opposite direction gave the same results. The conditions chosen in Figs 2 and 3 are particularly insensitive to light scattering because measuring and reference wavelengths are closely spaced. Further controls showed that the absorption changes with ferrocyanide were blocked by cyanide, uncoupling agents¹ or nigericin, and were drastically diminished in other conditions in which the mitochondrial pH gradient is minimized, for example when inorganic phosphate and succinate were added in the absence of *N*-ethylmaleimide.

Although it seems clear from these data, and from earlier data of others¹³⁻¹⁶, that the absorption changes of Neutral red measure changes in intramitochondrial H⁺ activity, the main conclusion is independent of this notion. As the measured absorption changes are fully energy-dependent, the results show in any case that cytochrome oxidase is twice as effective an energy converter as succinate-cytochrome c reductase. This is inconsistent with the hypothesis of Mitchell⁷ and Papa⁸⁻¹⁰, in which the energy-conserving efficiency is the same for these segments of the respiratory chain. It is also not consistent with the hypothesis of Azzone¹¹ and Lehninger *et al.*¹², according to which the energy-conserving efficiency is three times higher for the oxidase reaction than it is for succinate-cytochrome c reductase.

Conclusive evidence of membrane transport requires the demonstration of both uptake and release of the transported species on the two sides of the membrane. Measurements of proton release on the cytoplasmic side of the membrane and of translocation of electrical charges have led to the model shown in Fig. 1. The present demonstration of uptake of 2 H⁺/e⁻ from the matrix side of the membrane during cytochrome c oxidase activity completes this picture.

Whether Neutral red measures bulk matrix pH or more local H⁺-linked changes on the matrix side of the membrane cannot be decided on the basis of the present results. It can nevertheless be concluded that cytochrome oxidase functions as a proton pump with the proton-translocating stoichiometry suggested originally¹ (Fig. 1), and confirmed for cytochrome oxidase vesicles. The method presented here may also prove helpful in further assessment of proton translocation by enzymes other than cytochrome oxidase.

I thank Hilka Vuorenmaa for technical assistance and for help in preparation of the manuscript, and Matti Saraste for discussions and helpful suggestions. FCCP and myxothiazole were kind gifts of P. G. Heytler and H. Reichenbach, respectively. Research grants from the Sigrid Juselius Foundation, the Finnish Academy (MRC) and Finska Läkaresällskapet (Finnish Medical Society) are gratefully acknowledged.

Received 14 December 1983; accepted 15 February 1984.

- Wikström, M. *Nature* **266**, 271-273 (1977).
- Wikström, M. & Krab, K. *Biochim. biophys. Acta* **549**, 177-222 (1979).
- Casey, R. P., Chappell, J. B. & Azzi, A. *Biochem. J.* **182**, 149-156 (1979).
- Sigel, E. & Carafoli, E. *Eur. J. Biochem.* **111**, 299-306 (1980).
- Penttilä, T. *Eur. J. Biochem.* **133**, 355-361 (1983).
- Wikström, M., Krab, K. & Saraste, M. *A. Rev. Biochem.* **50**, 623-655 (1981).
- Mitchell, P. *Ann. N.Y. Acad. Sci.* **341**, 564-584 (1980).
- Lorusso, M., Capuano, F., Boffoli, D., Stefanelli, R. & Papa, S. *Biochem. J.* **182**, 133-147 (1979).
- Papa, S. *et al. Biochem. J.* **192**, 203-218 (1980).
- Papa, S., Guerrieri, F. & Izzo, G. *Biochem. J.* **216**, 259-272 (1983).
- Azzone, G. F., Pozzan, T. & Di Virgilio, F. *J. biol. Chem.* **254**, 10206-10212 (1979).
- Reynafarje, B., Alexandre, A., Davies, P. & Lehninger, A. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7218-7222 (1982).
- Junge, W., Ausländer, W., McGeer, A. & Runge, T. *Biochim. biophys. Acta* **546**, 121-141 (1979).
- Hong, Y. Q. & Junge, W. *Biochim. biophys. Acta* **722**, 197-208 (1983).
- Appelberg, B. G. & Saris, N.-E. *Abstr. 10th int. Congr. Biochem.*, Hamburg, 345 (International Union of Biochemistry, 1976).
- Pick, U. & Avron, M. *FEBS Lett.* **65**, 348-353 (1976).
- Mitchell, P., Moyle, J. & Mitchell, R. *Meth. Enzym.* **55**, 627-640 (1979).
- Wikström, M. & Saraste, M. in *New Comprehensive Biochemistry; Bioenergetics* (ed. Ernster, L.) (Elsevier, Amsterdam, in the press).
- Moyle, J. & Mitchell, P. *FEBS Lett.* **88**, 268-272 (1978).
- Wikström, M. & Krab, K. *FEBS Lett.* **91**, 8-14 (1978).

Errata

'Insight' in the pigeon: antecedents and determinants of an intelligent performance

R. Epstein, C. E. Kirshnit, R. P. Lanza & L. C. Rubin
Nature **308**, 61-62 (1984)

IN the third sentence of the first paragraph on page 61, 'a classic problem with that Köhler confronted his chimpanzees' should read 'a classic problem with which ...'.

VIP and noradrenaline act synergistically to increase cyclic AMP in cerebral cortex

Pierre J. Magistretti & Michel Schorderet
Nature **308**, 280-282 (1984)

ON page 281, the second and third paragraphs should read 'As shown in Fig. 1, the synergistic effect of 1 μM VIP and 10 μM NA was antagonized by the specific α-adrenergic antagonist phentolamine (100 μM) but not by (±)propranolol (100 μM), a specific β-adrenergic antagonist. Furthermore, as shown in Table 1, the specific α-adrenergic agonist phenylephrine ...'.

The mechanism of kainic acid neurotoxicity

CONCLUSIONS have been presented by Garthwaite and Garthwaite¹ based exclusively on light microscopic inspection of cerebellar slices incubated *in vitro* that the neurotoxic effects of kainic acid (KA) do not involve an interaction with L-glutamate released from excitatory terminals. They reject our proposal² that the KA-induced neuronal release of Glu and aspartate contributes to the neurotoxic action of KA. Furthermore, they suggest that the procedures used in these studies² were poorly designed.

The fundamental observation of Ferkany *et al.*² was that KA stimulates a calcium-dependent release of Glu and Asp from adult murine cerebellar slices incubated *in vitro*, an effect that has been extended to rat hippocampal and striatal slices³. The response is mediated by receptors specific for KA since neurophysiologically inactive analogues such as allo-KA and dihydro-KA are ineffective, as are also *N*-methyl-D,L-Asp and ibotenate, Glu analogues that act at different excitatory receptors. Tetrodotoxin does not block the effect of KA^{3,4}; thus, activation of dendritic KA receptors resulting in propagation of action potentials is not responsible for the release of Glu as proposed by Garthwaite and Garthwaite¹. Notably, 500 μ M KA is at least as effective in releasing Asp and Glu as is 35 mM K⁺, which curiously is considered "a moderate concentration of potassium ions" by the authors. Developmental and lesion studies indicate that the glutamatergic granule cells are the source of KA-induced release of Glu in cerebellum³. Our findings have been independently confirmed in slice studies on the lateral olfactory tract⁵, the striatum and the cerebral cortex⁶; in these latter studies, concentrations of KA ranging from 10 to 100 μ M have been shown to enhance the depolarization-induced release of Glu and D-Asp.

The authors imply that the results of our studies are artefactual since "the majority of cells in the slices used in this study were dead". Our slice technique has previously been demonstrated to maintain the tissue content of high-energy phosphates within the physiological range, the most reliable index of cell viability⁷. Using similar techniques, Nicklas *et al.*⁸ have demonstrated reversible effects of KA on ATP levels in cerebellar slices. As stated in our report, histological analysis of the slices did not reveal the severe disruption of cellular integrity described by Garthwaite *et al.*⁹ in mechanically chopped slices.

The primary argument proposed by these authors against the relevance of the presynaptic effects of KA on Glu release is that the response occurs at concentra-

tions of KA which "kill all neuronal cells in cerebellar slices". However, the authors have demonstrated previously that maximal stimulation of cyclic GMP and cyclic AMP formation in cerebellar slices occurs with 1–5 mM KA with dose-response curves resembling those described for KA-induced release of Glu and Asp¹⁰. The authors, furthermore, ignore the findings from tissue culture studies indicating that concentrations of KA ranging over 0.1–1.0 mM do not have short-term cytotoxic effects on isolated striatal explants devoid of glutamatergic innervation^{11,12} or on disaggregated cerebellar^{13,14} granule cells. And, as noted above, the presynaptic effect of KA in releasing Glu/Asp is detectable at 10 μ M in depolarizing conditions⁶, a concentration range deemed relevant to the neurotoxic action of KA by Garthwaite

Matters Arising

Matters Arising is meant as a vehicle for comment and discussion about papers that appear in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted. Neither contribution nor reply (if one is necessary) should be longer than 500 words and the briefest of replies, to the effect that a point is taken, should be considered.

and Garthwaite¹. Finally, their experiment demonstrating that the combination of 3 μ M KA and 3 mM Glu lacks neurotoxic effects cannot rule out the role of a presynaptic effect of KA since by their own criteria this is a subtoxic concentration of KA.

In conclusion, substantial evidence has accumulated that the neurotoxic effects of KA are complex and require the integrity of excitatory afferents in several systems examined^{15–17} and that KA directly stimulates or augments the depolarization-induced release of Glu/Asp from excitatory nerve terminals^{2–6}. The report of Garthwaite and Garthwaite¹ does not directly address either of these issues.

J. W. FERKANY
R. ZACZEK
J. T. COYLE*

Division of Child Psychiatry,
Departments of Psychiatry, Neuroscience,
Pharmacology and Pediatrics,
Johns Hopkins University School of
Medicine,
Baltimore, Maryland 21205, USA

* To whom reprint requests should be addressed.

1. Garthwaite, J. & Garthwaite, G. *Nature* **305**, 138–140 (1983).

2. Ferkany, J. W., Zaczek, R. & Coyle, J. T. *Nature* **198**, 757–759 (1982).
3. Ferkany, J. W. & Coyle, J. T. *J. Pharmac. exp. Ther.* **225**, 399–406 (1983).
4. Krespan, B., Berl, S. & Nicklas, W. J. *J. Neurochem.* **38**, 509–518 (1982).
5. Collins, G. G. S., Anson, J. & Surtees, L. *Brain Res.* **265**, 157–159 (1983).
6. Potashner, S. J. & Gerard, P. *J. Neurochem.* **40**, 1548–1557 (1983).
7. Biziere, K. & Coyle, J. T. *J. Neurochem.* **31**, 513–520 (1978).
8. Nicklas, W. J., Krespan, B. & Berl, S. *Eur. J. Pharmac.* **62**, 209–213 (1980).
9. Garthwaite, J., Woonhams, P. L., Collins, M. J. & Balazs, R. *Brain Res.* **173**, 373–377 (1979).
10. Garthwaite, J. *Neuroscience* **7**, 2491–2497 (1982).
11. Panula, P. A. *J. Brain Res.* **181**, 185–190 (1980).
12. Whetsell, W. O., Ecob-Johnston, M. S. & Nicklas, W. J. *Adv. Neurol.* **23**, 645–654 (1979).
13. Gibbs, W., Neale, E. A. & Moonen, G. *Dev. Brain Res.* **4**, 103–108 (1982).
14. Messer, A. & Maskin, P. *Neurosci. Lett.* **19**, 173–177 (1980).
15. Biziere, K. & Coyle, J. T. *J. Neurosci. Res.* **4**, 383–398 (1979).
16. Kohler, C., Schwarcz, R. & Fuxe, K. *Neurosci. Lett.* **10**, 241–246 (1978).
17. Streit, P., Stella, M. & Cuenod, M. *Brain Res.* **187**, 47–57 (1980).

GARTHWAITE AND GARTHWAITE
REPLY—In responding to our paper¹, Ferkany *et al.* review evidence that kainate can cause the release of endogenous excitatory amino acids from brain slices and that the neurotoxic effects of kainate in some brain areas or tissue culture systems are dependent on excitatory innervation being intact. Much of this information has been available for some years and these overall conclusions are not at issue.

In their paper², however, these authors took matters a stage further by claiming evidence for a causal link between the glutamate-releasing and neurotoxic effects of kainate in the cerebellum. There were two key elements in their proposal, both of which formed the basis of our own experiments. The first was that the neurotoxic actions of kainate involved the combined activation of postsynaptic receptors for kainate and the parallel fibre transmitter (glutamate). Thus, from studies of cyclic GMP accumulation, Ferkany *et al.* claimed that inhibition of the postsynaptic action of the parallel fibre transmitter prevented kainate neurotoxicity². Using more direct methods, we found this conclusion to be spurious¹. Unfortunately, in interpreting their cyclic GMP data, the authors may well have been misled by a report³ purporting to demonstrate that cyclic GMP responses to kainate are potentiated in the presence of glutamate, a result which has subsequently been shown to be erroneous⁴.

We showed, furthermore, using combinations of kainate and concentrations of glutamate known to activate postsynaptic receptors, that the availability of glutamate postsynaptically was unlikely to be a factor which limited the neurotoxic potency of kainate. More detailed experiments of this nature (including developmental studies) support this conclusion⁴.

(These particular experiments were, therefore, not carried out to test the role of a presynaptic effect of kainate (see response by Ferkany *et al.* above).)

The second important element in the proposal of Ferkany *et al.* was that kainate induced the release of glutamate from presynaptic parallel fibre terminals (thereby making this amino acid available for the hypothetical postsynaptic/neurotoxic effects discussed above). However, the dose-response curve for glutamate release by kainate (which covers the range 500 μ M–10 mM)^{2,5} was found to be so far removed from the concentration range having neurotoxic effects (5–10 μ M for the cells innervated by parallel fibres) that a causative relationship was considered unlikely¹. (In their response, Ferkany *et al.*, for some reason, misquote our results on two counts, first regarding the concentrations of kainate required to kill all neuronal cells (which is 30 μ M, that is, well below the minimum glutamate-releasing concentration) and second, with respect to cyclic nucleotide formation: the dose-response curve for the cyclic GMP-stimulating effect of kainate covers the range 3 μ M (threshold) to 100 μ M

(maximum)⁶ and is thus very different from the curve describing kainate-induced release of amino acids.)

As a final point, Ferkany *et al.* take exception to our interpretation¹ that the poor cyclic GMP elevations registered in their slices reflect poor cellular preservation (see ref. 7). Assuming the comments made in their response to mean that the postsynaptic elements were, in fact, preserved, this raises additional questions regarding the site and mechanism of glutamate release in the presence of kainate. Within a few minutes of application of neurotoxic concentrations of kainate, neurones become irreversibly depolarized and their membranes cease to resist the free passage of ions⁸. We note that Ferkany *et al.* used 15-min exposures for monitoring amino acid release⁵. It might be considered unrealistic to expect to demonstrate a direct primary effect of kainate on presynaptic terminals when, at the same time, these terminals (together with other neuronal and glial elements) are being subjected to the drastic changes in their ionic, metabolic and electrical environment that would be expected to accompany the postsynaptic neurotoxic

action of kainate (which may itself be Ca^{2+} -dependent⁹). The use of synaptosomal preparations for investigating presynaptic kainate receptors would therefore seem more appropriate than slices. However, kainate does not, apparently, cause significant release of glutamate from such preparations¹⁰.

J. GARTHWAITE

G. GARTHWAITE

Department of Veterinary Physiology
and Pharmacology,
University of Liverpool,
Brownlow Hill, PO Box 147,
Liverpool L69 3BX, UK

1. Garthwaite, J. & Garthwaite, G. *Nature* **305**, 138–140 (1983).
2. Ferkany, J. W., Zaczec, R. & Coyle, J. T. *Nature* **198**, 757–759 (1982).
3. Foster, G. A. & Roberts, P. J. *Neurosci. Lett.* **23**, 67–70 (1981).
4. Garthwaite, J. & Gilligan, G. J. *Neuroscience* (in the press).
5. Ferkany, J. W. & Coyle, J. T. *J. Pharmac. exp. Ther.* **225**, 399–406 (1983).
6. Garthwaite, J. *Neuroscience* **7**, 2491–2497 (1982).
7. Garthwaite, J., Woodhams, P. L., Collins, M. J. & Balazs, R. *Brain Res.* **173**, 373–377 (1979).
8. Constanti, A., Connor, J. D., Galvan, M. & Nistri, A. *Brain Res.* **195**, 403–420 (1980).
9. Duce, I. R., Donaldson, P. L. & Usherwood, P. N. R. *Brain Res.* **263**, 77–87 (1983).
10. Krespan, B., Berl, S. & Nicklas, W. J. *J. Neurochem.* **38**, 509–518 (1982).

Sequence similarity between putative gene products of geminiviral DNAs

STANLEY AND GAY¹ recently reported the determination of the complete nucleotide sequences of the two circular DNA molecules (DNAs 1 and 2) that make up the genome of cassava latent virus (CLV), a member of the geminivirus group. Several open reading frames (ORFs) capable of encoding proteins of molecular weights greater than 10,000 were identified in both strands of DNAs 1 and 2. Interestingly, two ORFs which are capable of encoding proteins of similar sizes (molecular weights 30,100 and 29,200) are present at equivalent positions on DNAs 1 and 2. We have compared the amino acid sequences of these proteins by a computer-assisted method described previously² and found that they are closely

related (Fig. 1). Of all the positions compared in the aligned sequences, 43% are occupied by residues that are identical or chemically similar (gaps were considered as substitutions between dissimilar amino acids regardless of their length). Furthermore, a statistical test shows that the similarity between the aligned sequences is highly significant, with a probability of occurrence by chance of 4.3×10^{-6} . The presence of such extensive sequence similarity suggests that like the ORF of DNA 1, which is known to code for the coat protein of CLV (ref. 1 and J. Stanley, personal communication), the ORF of DNA 2 also encodes a functional protein. Although no such extensive sequence similarities were found for other ORFs,

the present result, together with the similarity in size of the two viral DNAs and the fact that they share a very similar non-coding segment in common¹, also suggests that the bipartite viral genome has evolved from a single common ancestral DNA.

REIKO KIKUNO

HIROYUKI TOH

HIDENORI HAYASHIDA

TAKASHI MIYATA

Department of Biology,
Faculty of Science,
Kyushu University,
Fukuoka 812, Japan

1. Stanley, J. & Gay, M. R. *Nature* **301**, 260–262 (1983).
2. Toh, H., Hayashida, H. & Miyata, T. *Nature* **305**, 827–829 (1983).

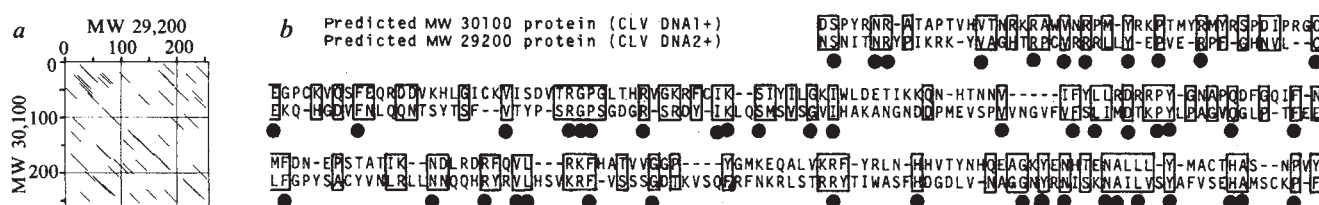


Fig. 1 *a*, Similarity matrix comparison of amino acid sequence of predicted 30,100 MW protein, encoded by the (+) strand of CLV DNA 1 (ordinate), with that of 29,200 MW protein encoded by the (+) strand of DNA 2 (abscissa). A computer program was used to generate diagonal lines indicating segments of 20 residues long that show sequence similarity (see ref. 2). *b*, Alignment of the 30,100 MW and 29,200 MW protein sequences. Aligned regions correspond to nucleotide positions 515–1,174 of DNA 1 and 632–1,327 of DNA 2 (see Fig. 1 of ref. 1) for 30,100 and 29,200 MW proteins, respectively. Boxed residues indicate identities (●) or favoured amino acid substitutions (see ref. 2). Gaps (—) were introduced to maximize similarity.

Science policy à la française

Robert Cahn

Simple Propos d'un Homme de Science. By Pierre Aigrain.

Hermann, 293 rue Lecourbe, 75015 Paris: 1983. Pp. 190. Pbk FF 74.

Pour une Prospective de la Science. Recherche et Technologie:

les Enjeux de l'Avenir. By Pierre Papon.

Seghers, 6 pl St-Sulpice, 75006 Paris: 1983. Pp. 383. Pbk FF 89.

SOME research scientists have the gift of popularizing the results of research, some research administrators have a penchant for analysing science policy and scientific institutions, some philosophers enjoy telling the practising scientist how he contrives to discover scientific truths, and rare politicians are able to say something unfamiliar about the ways in which scientific research may succeed in stimulating industrial developments. The two books under review, both of them by formerly practising scientists who have drifted into politics via science policy, are distinctly unusual because they each embrace all of these themes.

The two books are very alike yet very different. The themes treated by Aigrain and Papon largely overlap; the criticisms they make of the institutions and habits of their country, France, are uncannily similar. But their priorities and even ideologies are very different, and so their responses to the inadequacies they identify differ too.

In addition to the themes listed above, Papon focuses on two further issues — the role and nature of technological planning, and the problems of scientific forecasting. On the basis of many expert studies and of a number of early attempts which can be compared with the actual outcomes, he takes a sceptical view of forecasting, yet at the end of the book he outlines three alternative scenarios of the French scientific and technological future; these models, by his own admission, are impressionistic only, but none the worse for that, and they deserve close reading. As to the role of planning in the French scheme of things, this to Papon is absolutely central, the *sine qua non* of maintaining and advancing France's position among nations. His preferred scenario is predicated on rigid planning ("une mobilisation permanente et continue pendant deux décennies").

The differences in style, emphasis and values between the two authors, both of them physicists, reflect their different positions along the French political spectrum. Aigrain, who after the Second World War acquired a doctorate in semiconductor physics in the United States, returned to become a professor in Paris and helped to create the first solid-state physics research group in France; later he became research director of France's largest electronic enterprise, Thomson, and

moved on to become Minister of Research under Giscard d'Estaing. Papon, a younger man, worked as a research physicist in thermal physics in one of the laboratories of the Centre National de la Recherche Scientifique (CNRS) before becoming professor in one of the Parisian Grandes Ecoles; via membership of the "Committee of 10 Wise Men", which advised Aigrain, and a role as scientific counsellor of the re nascent Socialist Party, he became Director-General of the CNRS, his present post. Aigrain, the conservative, is quite sure that science and technology progressed most solidly during the years of de Gaulle's presidency — Pompidou, he admits, cared little for science. By contrast, Papon is critical of the long years of conservative administration; he is convinced that central definition of technological objectives, establishment of sharp priorities between different sciences, and detailed planning of both academic research and industrial development are crucial to France's emergence from the world economic crisis.

Both Aigrain and Papon are convinced internationalists, and are thoroughly familiar with "Anglo-Saxon" work in both science and science policy; Aigrain's American experience led him to foster Franco-American cooperation, whereas Papon is an Anglophile, at home with British science policy research and a wide range of books which he quotes in his remarkably varied bibliography. (Aigrain's book, which is based on a series of interviews and is much more informal, has no bibliography.)

The books have in common surveys of the early development of quantum physics, non-equilibrium thermodynamics, solid-state electronics and "informatics", and biotechnology, while Aigrain adds intriguing overviews of the present state of cosmology and of the brain in terms of current neurological models, and of Thom's catastrophe theory and its questionable extrapolation to sociology and ethology. Papon also discusses in some detail the interplay between nineteenth-century chemistry and the industrial competition between Germany and France. Both authors agree on the baleful influence of de Broglie, in the early years of quantum mechanics, in effectively suppressing French progress in this field because of his insistence on a deterministic view of the subject; to this, Papon adds the

comment that the opposition of Berthelot to the atomic hypothesis, together with Laplace's conviction that all science could be reduced to an analysis in terms of action-at-a-distance, delayed by decades the progress of nineteenth-century French chemistry.

These judgements on French scientific history strike the "Anglo-Saxon" reader as extraordinary: how can the convictions, however fierce, of a single scientist numb the progress of an entire scientific community? To appreciate this, it is necessary to have observed an old-fashioned French "patron" in action. All professional communities can be damaged by the powerful boss who always knows best, but in France the tyranny of the patron has in the past been in a class by itself. Woe betided the pupil or junior collaborator who sought a new post without the patron's advice and approval, or who sought to strike out in a new scientific direction! The most powerful patrons by degrees contrived to extend this stranglehold, initially over their immediate entourage, to the entire research communities in their fields. It is no wonder that the present administration of the CNRS, under Papon, has set an irrevocable limit to the years of authority of an individual head of a laboratory. The problem with this new rule is that many small laboratories (and most CNRS laboratories are very small by international standards) will lose their sense of direction when the patron is forced to hand over power to an inexperienced junior; the new CNRS rule, needed though it was, will work only where laboratories are combined into sizeable establishments which incorporate several lines of research, so that a number of scientists of similar standing work in a single administrative unit.

Both authors discuss with an air of desperation the rigidity of the French administrative machine. According to Papon, the modernization of the telephone system and the electrification of the railways were for decades wholly frustrated by interdepartmental rivalries. To quote Papon (in translation): "The major administrative machines, or the smaller services attached to individual Ministries, regard themselves to some degree as the holders of State power and of a measure of sovereignty itself . . .". One can only respond: "Yes, Minister!". (Yet de Gaulle, by Presidential strength, was able to force both these reforms through in a short time, and Giscard saw the admirable Train Grande Vitesse through to the verge of operation.)

Aigrain includes a fascinating chapter on his attempt to reform, in 1976, the Académie Française des Sciences, under Presidential instructions. The Academy, very influential in the early nineteenth century, had ossified into an affable club for eminent scientists in their seventies and

eighties: scientific categories had not altered in centuries (the term "biology" was not recognized before 1976) and the Government had not, it seems, asked the Academy for any official advice since 1917. Aigrain describes the very limited reforms, including an effective rejuvenation of the membership, he was able to push through against the concerted resistance of the Establishment. This chapter incorporates a critical comparison of Academies in several countries.

Another first-hand account by Aigrain deals with the early days of the Délégation Générale à la Recherche Scientifique et Technologique (DGRST), set up by de Gaulle in 1958. This was an interministerial body of scientific advisers, independent of the established civil service, which also had some research funds at its disposal. Its aim was to bypass the rigidity of the French Establishment, especially when the need arose to ease the way for new disciplines, such as molecular biology. There were few permanent staff, and most collaborators were scientists on a few years' part-time secondment. It is a source of some alarm that under the Mitterrand Government, this body (somewhat akin to the Science and Engineering Research Council in Britain) has been absorbed into the Research Ministry, has more permanent staff and is much more subject to the influence of centrally imposed technological objectives: the DGRST increasingly resembles the CNRS, itself modelled at its creation in 1938 on the Soviet Academy of Sciences. The upshot may be that technological-type planning is applied much too rigidly to pure science, and there are already clear signs of this. Aigrain's prediction (in translation) may prove apposite:

It is not impossible that fifty years from now, if the *mal français* persists, we shall end up with a Délégation Générale made up of a special corps of civil servants, graduates of a Grande Ecole created *ad hoc* and dug in for life.

The relationship between the DGRST and the CNRS emerges most interestingly when one reads the two books together; Papon in particular treats very frankly the difficulties of technological planning which both these bodies now attempt. This discussion is worth comparing with Sir Geoffrey Allen's analysis of the creeping intrusion of dirigisme in British science (*J. R. Soc. Arts* 129, 490; 1981).

Both authors, Papon in particular, also discuss the advantages and drawbacks of the French Grande Ecole system and its uneasy coexistence with the universities. Universities are bound to accept all school-leavers who have attained a minimum standard in the baccalauréat, the national school-leaving examination (though about half have to leave without completing their courses), while entry to the Grandes Ecoles, many of which have a mathematical bias, is only by success in a very highly competitive entrance examination for

which two years of intense full-time preparation is mandatory. Because of this, the Ecoles are largely middle-class preserves. However, the universities produce professionals such as doctors and lawyers. The university (French politicians realistically prefer to use the singular) is very closely and directly controlled by the Ministry of Education, whereas the Grandes Ecoles, under a variety of Ministries, have more independence and diversity. The Ecoles produce virtually all the senior civil servants and managing directors of large concerns in France. In spite of their manifold handicaps, the universities do most of the academic research. The Grandes Ecoles have, with some distinguished exceptions, proved durably

resistant to the implantation of research groups; yet they secure the best students and steer many of them towards industrial administration. French industry cannot, it seems, do without them, and on this the two authors, however different their viewpoints, appear to agree.

Aigrain and Papon are subject like all of us to the pull of ideology, yet they present their subjects honestly, warts and beauty-spots included. Both of these books are well worth reading, but the two together are more than their sum. □

Robert Cahn was formerly Professor of Materials Science at the University of Sussex and Professor of Metallurgy at the Université de Paris-Sud.

Utility of scarcity

R.K. O'Nions

Noble Gas Geochemistry.

By M. Ozima and

F.A. Podosek.

Cambridge University Press: 1983.

Pp.367. £40, \$79.50.

THE chemically inert character of the noble gases (sometimes known as rare or inert gases), together with their very low abundance in planetary objects, renders them particularly valuable natural tracers for both terrestrial and extraterrestrial processes. It is not surprising, therefore, that considerations of the elemental and isotopic abundances of noble gases are becoming increasingly prominent in geochemical research, as analytical techniques improve.

In recent years the existence of primordial volatiles within the Earth's mantle has been demonstrated using noble gas tracers, and important constraints have been placed on interpretations of the internal structure of the Earth. With the exception of the popular use of argon isotopes in conjunction with potassium as a geochronometer, the number of practitioners of noble gas geochemistry has been small. One can confidently predict, however, that this research area will expand significantly in the coming years and play a more central role in the earth sciences. The publication of *Noble Gas Geochemistry*, the first book devoted to a comprehensive discussion of this topic, is thus particularly timely.

In preparing a text in a rapidly evolving, specialized subject area, there is the inevitable risk that the book will become redundant shortly after publication. The emphasis adopted by Ozima and Podosek in *Noble Gas Geochemistry* ensures that this fate will be avoided. The book is likely to remain a highly relevant introductory text for years to come.

The authors start at the ground floor, and in the first four chapters present the

relevant nomenclature and the basic physical and nuclear chemistry of the noble gases. Although the subject matter of the first chapters necessarily contains a considerable amount of tabulated data, it is rendered palatable by a lucid and easy style of presentation: there is here a great deal of useful, accessible reference data. The same lucid style characterizes the remaining chapters, which systematically deal with the abundance and isotopic compositions of the noble gases in meteorites and in the Earth's hydrosphere, mantle and igneous and sedimentary systems. The layout is somewhat encyclopaedic in style, but this is in fact a positive aid to finding information rather than evidence of pedantry on the part of the authors. These sections are well written, packed with valuable information, and reflect the authors' exceptionally good grasp of the subject material. The last chapter reviews the models of Earth degassing and atmospheric evolution that have been proposed on the basis of the abundances and isotopic composition of noble gases.

If the hallmark of the book is the clarity of presentation — extending beyond the text, with its many poignant comments, to the well-prepared tables and figures — its strongest point is the thorough discussion of the basic principles. At this time the book is well up to date, with an excellent bibliography that includes references to many of the recent developments.

Noble Gas Geochemistry is likely to be invaluable to geochemists and will find wide use as an introduction to the subject for a range of non-specialists at both the undergraduate and graduate level. While it is not cheap, I suspect that many potential readers will have to buy it because the library copy will always be out on loan. Ozima and Podosek have performed a commendable task in producing a text on what they themselves term "the utility of scarcity". □

R.K. O'Nions is a Royal Society Professor in the Department of Earth Sciences, University of Cambridge.

Comparatively better

Joe Felsenstein

The Explanation of Organic Diversity: The Comparative Method and Adaptions for Mating.

By Mark Ridley.

Clarendon: 1983. Pp.272. £19, \$37.50.

THE use of cross-species comparisons to evaluate evolutionary patterns has been growing rapidly in recent years. And, as being explicit and numerical has become more respectable among systematists, interest in understanding the logical and statistical pitfalls of the comparative method has also increased. In *The Explanation of Organic Diversity*, Mark Ridley outlines a comparative method that is intended to take correct account of phylogeny. He illustrates its use by examining the distribution of two phenomena: precopulatory mate-guarding in arthropods and anurans, and assortative mating for size in a variety of phyla.

The first 41 pages of the monograph will attract the widest attention. Here Ridley presents a critique of previous approaches and an explanation of his own method. Previous workers have often engaged in statistical tests of association between a phenotype and an environment, using indi-

vidual species as the data points. The problem with this practice is that the statistical tests presume that the species are independent, whereas they in fact are found in non-independent clusters in a hierarchical phylogeny. This clustering creates meaningless correlations: we may find that all bird species living on mountaintops have long tails, but fail to notice that they are a group of close relatives, and that the long tails were acquired only once, in their common ancestor.

Ridley joins the ranks of those — such as Clutton-Brock, Harvey, Baker and Parker — who are alarmed by the possibilities for statistical bias in using species as if they were independent entities. The method he proposes involves finding or constructing a phylogeny for the group under study and then, by minimizing the number of times that an independent origin must be assumed, inferring those points at which the phenotype has arisen independently. These independent occurrences are then taken as the individual instances in constructing a contingency table relating the phenotype to the state of the environment.

Although this method, also previously used by John Gittleman, is immeasurably superior to previous practice, it does not completely exorcise the statistical problems. In assigning positions of changes by minimizing the amount of change (a criterion that has acquired the

unfortunate name of ‘parsimony’, though Ridley mercifully avoids the term), one does not achieve an error-free assignment. If, for example, two sister lineages A and B both have their phenotype in a derived state P and both exist in environment E, we minimize the number of independent origins of P by assuming that it arose once, in the ancestor of A and B, which we assume also to have been in environment E. In truth, P may have arisen separately in the two lineages which themselves may or may not have entered environment E separately. The placements of changes in the phylogeny are thus subject to error, so that there is more statistical uncertainty than Ridley’s test implies. I cannot persuade myself that the test is always conservative. In treating the inferences of placements of changes as if they were observations, Ridley’s test does not entirely achieve statistical respectability even though it is a considerable step in that direction.

The monograph is written with clarity, considerable wit and a lack of false modesty. Without muddying the waters, Ridley is able to refute most of the arguments of those who criticize the comparative method. I was particularly relieved to see that, although declaring his method to be ‘cladism’, he was able clearly to distinguish classification from phylogeny, something few systematists do. Cladism, which is always being confused with completely unrelated doctrines such as punctualism or even Marxism, is genuinely ambiguous on one point: whether it is a position on classification or a set of methods for inferring evolutionary history. Ridley uses Hennig’s methods in the latter sense, explicitly disavowing any intention of taking a position on methods of classification. His dismissal of phylogenetic inertia is much less satisfying, however. The mysterious declaration that his methods could not cope with this inertia, coupled with his apparent lack of concern over such a state of affairs, leave me with the impression that I am in the presence of the truest of panslectionists.

The bulk of the monograph is a detailed examination of precopula and of assortative mating by size. For each of these a prediction is tested, using his methods, and the data and method are presented clearly. Ridley is frequently forced to use existing classifications as if they were phylogenies, but unlike many authors he is aware of the limitations of this practice.

If Ridley’s monograph is able to persuade comparative biologists to abandon counting species as though they represented independent evolutionary events, it will have done noble work. But it would be unfortunate if Ridley’s clarity and persuasiveness caused his methods to be taken as the last word on the statistical analysis of comparative data. □

Joe Felsenstein is a Professor in the Department of Genetics, University of Washington, Seattle.

ADVERTISEMENT

New from Princeton

Catastrophes and Earth History

The New Uniformitarianism
W.A. Berggren and
John A. Van Couvering
Princeton Series in Geology
and Paleontology
Alfred G. Fischer, Editor
C \$65.00. P \$19.50

Earth's Earliest Biosphere

Its Origin and Evolution
Edited by J. William Schopf
C \$95.00. P \$42.50
Winner of the 1983 Professional and
Scholarly Publishing Award
Physical Sciences
Association of American Publishers

The History of the Earth's Crust

A Symposium
Edited by Robert A. Phinney
C \$33.00. P \$11.50

Quantum Theory and Measurement

Edited by John Archibald Wheeler
and Wojciech Hubert Zurek
Princeton Series in Physics
P.W. Anderson and
A.S. Wightman, Editors
C \$75.00. P \$19.50

Random Walks in Biology

Howard C. Berg
\$16.50
Co-winner of the 1984 Biological
Physics Prize of the American
Physical Society

Molecular Approaches to Evolution

Jacques Ninio
Translated by Robert Lang
C \$25.00. P \$12.50

Mate Choice in Plants

Mary F. Willson and Nancy Burley
Monographs in Population
Biology, 19
Robert M. May, Editor
C \$35.00. P \$12.50

Kinetics of Heterogeneous Catalytic Reactions

Michel Boudart and G. Djéga-Mariadassou
Physical Chemistry: Science and
Engineering
John M. Praunitz and Leo Brewer,
Editors

Princeton
University Press

41 William Street Princeton, NJ 08540



Routes to living on the land

James D. Robertson

The Colonisation of Land: Origins and Adaptations of Terrestrial Animals.

By Colin Little.

Cambridge University Press: 1983.

Pp.290. £55, \$99.50.

TO A biologist, the colonization of land by animals first conjures up a simple progressive sequence — lungfish-like creatures evolving into amphibians, some of which became better adapted to drier conditions as reptiles, and from the latter group a line arising which led to the mammals. But such a picture ignores the many terrestrial invertebrates. The chordate groups constitute only one of about thirty phyla, and nine of the remaining twenty-nine invertebrate phyla contain members living on land, the most abundant being arthropods, gastropod molluscs and earthworms.

Dr Little has undertaken the task of evaluating the literature on the probable evolution of the terrestrial members of each phylum from an aquatic medium, and the much greater volume of work available on the morphological and physiological adaptations to the new environment, especially in relation to salt and water balance (including permeability, nitrogenous excretion, respiration and reproduction). Colonization of land from the sea has four possible routes — directly via the marine littoral zone; via mangroves and the salt-marsh habitat; from brackish and fresh water; or through the subterranean interstitial ground water between sea, fresh water and land. Then, of course, there is the reverse movement of terrestrially adapted animals back to an aquatic environment.

Animals on land vary in the degree to which they are completely emancipated from water. Thus different families of land-crabs seem to have reached land via the marine littoral or brackish water route; crabs have to go back to the sea to breed, but their gill filaments do not collapse when the crabs are out of water and there is usually increased vascularization of the gill chambers with a thinning of the cuticle in the upper parts to form lungs. By contrast, other land crustaceans such as isopods and amphipods are considered to have taken the marine littoral route without undergoing substantial structural and physiological changes. The same path seems also to have been used by nemertines and chelicerates, the latter including the arachnids, scorpions and other groups each of which may have arisen at different times from marine ancestors.

Little argues that the interstitial environment in both marine and freshwater sands, and in marine and terrestrial ground water, has been the route to terrestrial soils of

small animals such as protozoans, nematodes and ostracods. There is, he says, even the possibility that the ancestors of myriapods and insects came from a marine interstitial habitat.

From fresh water seem to have come platyhelminths, earthworms and leeches, crayfishes and the vertebrates. Terrestrial molluscs appear to have used three routes — two prosobranch groups arriving via fresh water; another prosobranch group (a family related to the marine intertidal *Littorina* spp.) via the marine littoral and possibly the mangrove habitat; and the pulmonates from salt marshes.

All the conclusions reached by Little are based on the evidence of present distribution and, especially, on considerations of comparative morphological, behavioural and physiological data, including modifications in excretory organs, integuments, respiratory surfaces, the osmotic and ionic composition of haemolymph and the characteristics of its

respiratory pigments, type of nitrogenous excretion, and other features. Indeed, for many the greatest value of the book will be the up-to-date information on the ecological physiology of aquatic and terrestrial groups. The data are given in 49 tables and 132 figures, all of which (apart from the 26 photographs) are redrawn from published work in an eminently clear style. There are 1,525 references and adequate general, taxonomic and author indexes.

As the culmination of ten years' work, Dr Little has produced a valuable book which is written with great clarity. He has brought together an immense amount of information on adaptations to aquatic and terrestrial habitats, particularly in the invertebrate groups, and has put forward reasonable conclusions on the origins of different terrestrial animals from an aquatic habitat. □

James D. Robertson is at the Department of Zoology, University of Glasgow.

Planning prehistory

Robert Foley

Ascent to Civilization: The Archaeology of Early Man.

By John Gowlett.

Collins/Knopf: 1984. Pp.208.

£10.95; pbk \$14.95.

THE character of our own species can be illuminated from a variety of sources. Comparative anatomy, the behaviour of primates, fossil evidence, the pattern of contemporary humans and of their cultures can and have all been used to tell the story of the emergence of man. Books on the subject, stretching back to Lyell, Lubbock and Darwin, have ranged from the dogmatic to the doggedly detailed to the downright lunatic. The latest, by John Gowlett, is based firmly on the archaeological record, concentrates on the material evidence rather than speculation, analogy or theory, and certainly has the advantage over its nineteenth-century forerunners of excellent photographs and diagrams of the sites of early man.

From its title, *Ascent to Civilization* ought to be extracted highlights from two television series — Jacob Bronowski's *The Ascent of Man* and Kenneth Clark's *Civilization*. However, where Bronowski and Clark dwelt on the last two thousand years, Gowlett is interested in the preceding two million, the achievements of which are written in stone and bone. Despite this chronological discrepancy, the book is comparable to those two series in terms of clarity, lavishness and the tracing of cultural development through material products.

Gowlett's task is to document the character of man from the fossil and archaeological evidence of the Stone Age,

but the context is provided by discussions of the relationship between hominids and apes, by theories of human evolution, and especially by sections showing the nature of archaeological and palaeoenvironmental methodology. In this, two points stand out. First Gowlett is keen to present the basic evidence of archaeology, not to pursue the labyrinthine history of discovery and terminology. And, second, he is remarkably up-to-date and accurate across an enormous range of time and material. Consequently the book provides an extremely accessible and perceptive account of the Stone Age.

Although Gowlett's principal aim is to show the pattern of prehistory, his own conception of what it means to be human comes through. Man is a planner. His success and uniqueness as a species reflects his ability for forethought, and prehistory records increasing planning and organization of human societies. This emphasis on mental abilities marks a return to earlier views on human evolution, in contrast to the currently popular bioenergetic and mechanistic approaches. Gowlett makes a good case for seeing the human brain as more than an environmental epiphenomenon, and for tracking mental abilities through the archaeological record. But the doubt remains that this is explanation by hindsight. The benefits of planning and forethought seem self-evident now, but the reasons why (despite the costs involved) they should have been selected for in a primate in Africa some two million years ago remain to be revealed. However, in this excellent book Gowlett has made available to a wide audience the wealth of material and methods on which the debate about early man must be based. □

Robert Foley is a Lecturer in the Department of Anthropology at the University of Durham.

Mathematics

Banach Bundles, Banach Modules and Automorphisms of C^* -Algebras. Research Notes in Mathematics, Vol. 92. By M.J. DUPRÉ and R.M. GILLETTE. Pitman: 1983. Pp. 11. Pbk ISBN 0-273-08626-X. Np.

Birkhoff Interpolation. Encyclopedia of Mathematics and Its Applications, Vol. 19. Addison-Wesley: 1983. Pp. 237. ISBN 0-201-13518-3. \$32.50.

Contributions to Nonlinear Partial Differential Equations. Research Notes in Mathematics, Vol. 89. By C. BARDOS *et al.* Pitman: 1983. Pp. 284. Pbk ISBN 0-273-08595-6. Np.

Encyclopedia of Statistical Sciences, Vol. 3. S. KOTZ and N.L. JOHNSON (eds). Wiley: 1983. Pp. 722. ISBN 0-471-05549-2. £75.50, \$105.75.

The Method of Differential Approximation. By Y.I. SHOKIN. Springer-Verlag: 1983. Pp. 296. ISBN 3-540-12225-7/0-387-12225-7. DM 96, \$41.40.

Nonlinear Partial Differential Equations and Their Applications. Collège de France Seminar, Vol. 5. Research Notes in Mathematics, Vol. 93. By H. BREZIS and J.L. LIONS. Pitman: 1983. Pp. 370. Pbk ISBN 0-273-08620-0. Np.

Similarity Solutions of Nonlinear Partial Differential Equations. Research Notes in Mathematics, Vol. 88. By L. DRESNER. Pitman: 1983. Pp. 124. Pbk ISBN 0-273-08621-9. Np.

Astronomy

The A-Stars: Problems and Perspectives. By S.C. WOLFF. National Aeronautics and Space Administration: 1983. Pp. 211. Np.

Astronomy and History: Selected Essays. By O. NEUGEBAUER. Springer-Verlag: 1983. Pp. 538. Pbk ISBN 0-387-90844-7/3-540-90844-7. DM 56, \$21.80.

Astrophysics and Space Physics Reviews, Vol. 2. R.A. SYUNYAEV (ed.). Harwood: 1983. Pp. 449. ISBN 3-7186-0070-6. \$185.

The Atlas of the Solar System. By P. MOORE *et al.* Mitchell Beazley: 1983. Pp. 464. ISBN 0-85533-468-1. £19.95.

An Atlas of UV Spectra of Supernovae. European Space Agency: 1983. Pp. 22. FF 140.

Cosmic Ecology: The View From The Outside In. By G.A. SEIELSTAD. University of California Press: 1983. Pp. 174. ISBN 0-520-04753-2. \$24.95.

Developing the Space Frontier, Vol. 52: Advances in the Astronautical Sciences. A. NAUMANN and G. ALEXANDER (eds). American Astronautical Society: 1983. Pp. 417. Hbk ISBN 0-87703-184-3; pbk ISBN 0-87703-185-1. Hbk \$55; pbk \$45.

Encyclopedia of Astronomy. S.P. PARKER (ed.). McGraw-Hill: 1983. Pp. 450. ISBN 0-07-045251-2. £36.25.

Space Manufacturing 1983. Vol. 53: Advances in the Astronautical Sciences. J.D. BURKE and A.S. WHITT (eds). American Astronautical Society: 1983. Pp. 478. Hbk ISBN 0-87703-188-6; pbk ISBN 0-87703-189-4. Hbk \$60; pbk \$50.

Supernova Remnants and Their X-Ray Emission. J. DANZIGER and P. GORENSTEIN (eds). Reidel: 1983. Pp. 614. Hbk ISBN 90-277-1666-8; pbk ISBN 90-277-1667-6. Hbk Dfl. 180, \$72; pbk Dfl. 95, \$38.

Understanding the Universe: The Impact of Space Astronomy. R.M. WEST (ed.). Reidel: 1983. Pp. 249. ISBN 90-277-1647-1. Dfl. 90, \$34.50.

Chemistry

Circular Dichroic Spectroscopy: Exciton Coupling in Organic Stereochemistry. By N. HARADA and K. NAKANISHI. University Science Books: 1983. Pp. 460. ISBN 0-935702-09-1. \$32.

Introduction to Polymer Viscoelasticity, 2nd Edn. By J.J. AKLONIS and W.J. MACKNIGHT. Wiley: 1983. Pp. 295. ISBN 0-471-86729-2. £37.95, \$53.

Organic Syntheses with Carbon-14. By R.R. MUCCINO. Wiley: 1983. Pp. 676. ISBN 0-471-05165-9. £49.95, \$69.95.

Polarographic Oxygen Sensors: Aquatic and Physiological Applications. E. GNAIGER and H. FORSTNER (eds). Springer-Verlag: 1983. Pp. 370. ISBN 3-540-11654-0/0-387-11654-0. DM 118, \$50.90.

Radiation Chemistry, Vols 1 and 2. Proceedings of the 5th Tihany Symposium. By J. DOBÓ, P. HEDVIG and R. SCHILLER. Akadémiai Kiadó: 1983. Pp. 1142. Vol. 1 ISBN 963-05-3430-4; Vol. 2 ISBN 963-05-3431-2. Vol. 1 \$75; Vol. 2 \$75.

Physics

Advances in Applied Mechanics, Vol. 23. J.W. HUTCHINSON and T.Y. WU (eds). Academic: 1983. Pp. 453. ISBN 0-12-002023-8. \$74.

Advances in Magnetic Resonance, Vol. 11. J.S. WAUGH (ed.). Academic: 1983. Pp. 282. ISBN 0-12-025511-1. \$49.50.

Atomic and Molecular Physics of Controlled Thermonuclear Fusion. NATO Advanced Science Institutes Series, Vol. 101. C.J. JOACHAIN and D.E. POST (eds). Plenum: 1983. Pp. 576. ISBN 0-306-41398-1. \$79.50.

Atomic Physics, Vol. 8. I. LINDGREN, A. ROSEN and S. SVANBERG (eds). Plenum: 1983. Pp. 592. ISBN 0-306-41361-2. \$79.50.

Beam Optics of Quadrupole Probe-Forming Systems. By G.W. GRIME and F. WATT. Adam Hilger: 1983. Pp. 263. ISBN 0-85274-546-X. £27.50.

Brewster and Wheatstone on Vision. N.J. WADE (ed.). Academic: 1983. Pp. 360. ISBN 0-12-729550-X. £25, \$39.

Cryocoolers, Part 1: Fundamentals. By G. WALKER. Plenum: 1983. Pp. 365. ISBN 0-306-40715-9. \$49.50.

Electro-Optical and Magneto-Optical Properties of Liquid Crystals. By L.M. BLINOV. Wiley: 1983. Pp. 341. ISBN 0-471-10231-8. £35.95, \$69.95.

Fundamental Processes in Energetic Atomic Collisions. NATO Advanced Science Institutes Series, Vol. 103. H.O. LUTZ, J.S. BRIGGS and H. KLEINPOPPEN (eds). Plenum: 1983. Pp. 675. ISBN 0-306-41465-1. \$95.

Group Theory in Spectroscopy: With Applications to Magnetic Circular Dichroism. By S.B. PIEPHO and P.N. SCHATZ. Wiley: 1983. Pp. 0-471-03302-2. £65.75, \$91.95.

Interferometry, 2nd Edn. By W.H. STEEL. Cambridge University Press: 1983. Pp. 308. ISBN 0-521-25320-9. £35, \$69.50.

An Introduction to Solar Radiation. By M. IQBAL. Academic: 1983. Pp. 390. Hbk ISBN 0-12-373750-8; pbk ISBN 0-12-373752-4. Hbk np; pbk \$32.

Modern Logic and Quantum Mechanics. By R.W. GARDEN. Adam Hilger: 1983. Pp. 178. ISBN 0-85274-457-9. £18.

Nonlinear Dynamics and Turbulence. G.I. BARENBLATT, G. LOOSS and D.D. JOSEPH (eds). Pitman: 1983. Pp. 356. ISBN 0-273-08560-3. £35.

The Quantum Mechanical Few-Body Problem. By W. GLÖCKLE. Springer-Verlag: 1983. Pp. 197. ISBN 3-540-12587-6/0-387-12587-6. DM 82, \$31.90.

Quantum Metrology and Fundamental Physical Constants. NATO Advanced Science Institutes Series, Vol. 98. P.H. CUTLER and A.A. LUCAS (eds). Plenum: 1983. Pp. 658. ISBN 0-306-41372-8. \$95.

Semiconductors and Semimetals, Vol. 19: Deep Levels, GaAs, Alloys, Photochemistry. R.K. WILLARDSON and A.C. BEER (eds). Academic: 1983. Pp. 342. ISBN 0-12-752119-4. \$57.50.

Semi-Riemannian Geometry: With Applications to Relativity. By B. O'NEILL. Academic: 1983. Pp. 468. ISBN 0-12-526740-1. \$45.

A Sophisticate's Primer of Relativity, 2nd Edn. By P.W. BRIDGMAN. Wesleyan University Press: 1983. Pp. 172. Pbk ISBN 0-8195-6078-2. \$8.95.

Sources and Applications of Ultraviolet Radiation. By R. PHILLIPS. Academic: 1983. Pp. 440. ISBN 0-12-553880-4. £35, \$60.

Sputtering by Particle Bombardment. II: Sputtering of Alloys and Compounds, Electron and Neutron Sputtering Surface Topography. R. BEHRISCH (ed.). Springer-Verlag: 1983. Pp. 391. ISBN 3-540-12593-0/0-387-12593-0. DM 109, \$42.30.

Superconducting Rotating Electrical Machines. By J.R. BUMBY. Oxford University Press: 1983. Pp. 173. ISBN 0-19-859327-9. £25.

Theoretical Mechanics. By E.N. MOORE. Wiley: 1983. Pp. 456. ISBN 0-471-87488-4. £24.75, \$34.95.

Time-Resolved Vibrational Spectroscopy. G.H. ATKINSON (ed.). Academic: 1983. Pp. 398. ISBN 0-12-066280-9. \$43.

Topological Disorder in Condensed Matter. Springer Series in Solid-State Sciences, Vol. 46. F. YONEZAWA and T. NINOMIYA (eds). Springer-Verlag: 1983. Pp. 253. ISBN 3-540-12663-5/0-387-12663-5. DM 70, \$27.20.

Earth Sciences

Arc Volcanism. Developments in Volcanology, Vol. 2. S. ARAMAKI and I. KUSHIRO (eds). Elsevier Biomedical/North-Holland: 1983. Pp. 652. ISBN 0-444-42234-X. \$97.75, Dfl. 230.

Drilling Discharges in the Marine Environment. National Academy of Sciences Press: 1983. Pp. 180. Pbk ISBN 0-309-03431-0. \$14.75.

Explosive Volcanism. Developments in Volcanology, Vol. 3. M.F. SHERIDAN and F. BARBERI (eds). Elsevier Biomedical/North-Holland: 1983. Pp. 481. ISBN 0-444-42251-X. \$84.75, Dfl. 220.

Glacial-Marine Sedimentation. B.F. MOLNIA (ed.). Plenum: 1983. Pp. 844. ISBN 0-306-41497-X. \$65.

Intracontinental Fold Belts: Case Studies in the Variscan Belt of Europe and the Damara Belt in Namibia. H. MARTIN and F.W. EDER (eds). Springer-Verlag: 1983. Pp. 945. ISBN 3-540-12440-3/0-387-12440-3. DM 120, \$47.70.

Introductory Dynamical Oceanography, 2nd Edn. By S. POND and G.L. PICKARD. Pergamon: 1983. Pp. 329. Hbk ISBN 0-08-028729-8; pbk ISBN 0-08-0287-28-X. Hbk np; pbk £6.95, \$12.50.

Mineral Deposits of the Alps and of the Alpine Epoch in Europe. H.-J. SCHNEIDER (ed.). Springer-Verlag: 1983. Pp. 402. ISBN 3-540-12231-1/0-387-12231-1. DM 98, \$39.

The Phanerozoic Geology of the World. II: The Mesozoic. B.M. MOULLADE and A.E.M. NAIRN (eds). Elsevier Biomedical/North-Holland: 1983. Pp. 450. ISBN 0-444-41672-2. \$112.75, Dfl. 265.

Plateau Uplift: The Rhenish Shield — A Case History. K. FUCHS *et al.* (eds). Springer-Verlag: 1983. Pp. 411. ISBN 3-540-12577-9/0-387-12577-9. DM 104, \$40.40.

Practical Sedimentology. By D.W. LEWIS. Hutchinson Ross: 1984. Pp. 227. Pbk ISBN 0-87933-443-6. Np.

Synthesis and Modelling of Intermittent Estuaries: A Case Study for Planning to Evaluation. Lecture Notes on Coastal and Estuarine Studies, Vol. 3. W.R. CUFF and M. TOMCZAK Jr (eds). Springer-Verlag: 1983. Pp. 302. Pbk ISBN 3-540-12681-3/0-387-12681-3. DM 49.80, \$19.40.

Wandering Continents and Spreading Sea Floors on an Expanding Earth. By L.C. KING. Wiley: 1983. Pp. 232. ISBN 0-471-90156-3. £15.75, \$29.95.

Technology

The Acetone-Butanol Fermentation and Related Topics 1980-83. J.D. BU'LOCK and A.J. BU'LOCK (eds). Science and Technology Letters, Kew, Surrey, UK: 1983. Pp. 150. ISBN 0-946682-00-9. Np.

Austenitic Steels at Low Temperatures. R.P. REED and T. HORIUCHI (eds). Plenum: 1983. Pp. 388. ISBN 0-306-41371-X. \$62.50.

Coal Liquefaction Products, Vol. 1: NMR Spectroscopic Characterization and Production Processes. H.D. SCHULTZ (ed.). Wiley: 1983. Pp. 415. ISBN 0-471-89232-7. £53.30, \$74.95.

Deep Space Telecommunications Systems Engineering. J.H. YUEN (ed.). Plenum: 1983. Pp. 603. ISBN 0-3-06-41489-9. \$55.

Glass: Science and Technology, Vol. 1: Glass-Forming Systems. D.R. UHLMANN and N.J. KREIDL (eds). Academic: 1983. Pp. 465. ISBN 0-12-706701-9. \$74.50.

Imaging With Non-Ionizing Radiations, Vol.2. D.F. JACKSON (ed.). Surrey University Press: 1983. Pp.228. ISBN 0-903384-37-X. £26.50.

Noise In Electronic Devices and Systems. By M.J. BUCKINGHAM. Ellis Horwood/Halsted: 1983. Pp.372. ISBN 0-85312-218-0/0-470-27467-0. £35, \$60.50.

Power Lasers and Their Applications. By V.S. LETOKHOV and N.D. USTINOV. Harwood: 1983. Pp.128. ISBN 3-7186-0166-4. \$42.50.

PowerSources 9: Research and Development in Non-Mechanical Electrical Power Sources. Proceedings of the 13th International Symposium held September 1982, Brighton. J. THOMPSON (ed.). Academic: 1983. Pp.592. ISBN 0-12-689160-5. £66, \$99.50.

Quantitative Electron-Probe Microanalysis. V.D. SCOTT and G. LOVE (eds). Ellis Horwood/Halsted: 1983. Pp.345. Hbk ISBN 0-85312-514-7; pbk ISBN 0-85312-672-0. Hbk np; pbk £12.50, \$21.65.

Remote Sensing: New Satellite Systems and Potential Applications. European Space Agency: 1983. Pp.166. FF 90.

Space Applications at the Crossroads. Science and Technology Series, Vol.55. J.H. McELROY and E.L. HEACOCK (eds). American Astronautical Society: 1983. Pp.296. Hbk ISBN 0-87703-186-X; pbk ISBN 0-87703-187-8. Hbk \$45; pbk \$35.

Techniques for Image Processing and Classification in Remote Sensing. By R.W. SCHOWENGERDT. Academic: 1983. Pp.249. ISBN 0-12-628980-8. \$25.

Treatise on Materials Science and Technology, Vol.25: Embrittlement of Engineering Alloys. C.L. BRIANT and S.K. BANERJI (eds). Academic: 1983. Pp.623. ISBN 0-12-341825-9. \$89.

The Trouble with Technology: Explorations in the Process of Technological Change. S. MACDONALD, D. McL. LAMBERTON and T. MANDEVILLE (eds). Frances Pinter: 1983. Pp.224. ISBN 0-86187-285-1. £16.

Wave Propagation and Remote Sensing. Proceedings of URSI Commission F Symposium held June 1983, Belgium. European Space Agency: 1983. Pp.493. No Pbk ISBN. FF 175.

Biological Sciences

Coherent Excitations in Biological Systems. H. FRÖHLICH and F. KREMER (eds). Springer-Verlag: 1983. Pp.224. ISBN 3-540-12540-X/0-387-12540-X. DM 59, \$24.40.

Contemporary Nephrology, Vol.2. S. KLAHR and S.G. MASSRY (eds). Plenum: 1983. Pp.793. ISBN 0-306-41303-5. \$75.

Culture of Marine Invertebrates: Selected Readings. C.J. BERG Jr (ed.). Hutchinson Ross: 1983. Pp.385. ISBN 0-87933-105-4. Np.

Current Topics in Microbiology and Immunology, Vol.105. M. COOPER et al. (eds). Springer-Verlag: 1983. Pp.184. ISBN 3-540-12492-6/0-387-12492-6. DM 112, \$43.50.

Development in Mammals, Vol.5. M.H. JOHNSON (ed.). Elsevier Biomedical/North-Holland: 1983. Pp.381. ISBN 0-444-80506-0. \$93, Dfl. 220.

Development of the Vascular System. Proceedings of a Ciba Foundation Symposium held January 1983. Pitman: 1983. Pp.254. ISBN 0-272-79732-4. £25, \$35.

The Diversity of Crop Plants. By J.G. HAWKES. Harvard University Press: 1983. Pp.184. ISBN 0-674-21286-X. £16.

Elm. By R.H. RICHENS. Cambridge University Press: 1983. Pp.347. ISBN 0-521-24916-3. £35, \$69.50.

Endocrine System. T.C. JONES, U. MOHR and R.D. HUNT (eds). Springer-Verlag: 1983. Pp.366. ISBN 3-540-11677-X/0-387-11677-X. DM 160, \$63.60.

Environmental Adaptations. The Biology of Crustacea, Vol.8. F.J. VERNBERG and W.B. VERNBERG (eds). Academic: 1983. Pp.383. ISBN 0-12-106408-5. \$49.50.

Excited States of Biopolymers. R.F. STEINER (ed.). Plenum: 1983. Pp.258. ISBN 0-306-41316-7. \$39.50.

Gaseous Loss of Nitrogen from Plant-Soil Systems. Developments in Plant and Soil Sciences, Vol.9. J.R. FRENEY and J.R. SIMPSON (eds). Martinus Nijhoff: 1983. Pp.317. ISBN 90-247-2820-7. DG 150, \$65.50.

The Genetics and Biology of Drosophila, Vol.3c. M. ASHBURNER, H.L. CARSON and J.N. THOMPSON Jr (eds). Academic: 1983. Pp.484. ISBN 0-12-064947-0. £52, \$86.

The Geographical Distribution of Animal Viral Diseases. By S. ODEND'HAL. Academic: 1983. Pp.493. ISBN 0-12-524180-1. \$49.50.

Hepatocellular Carcinoma: A Series of Workshops on the Biology of Human Cancer Report No.17. K. OKUDA and I. MACKAY (eds). International Union Against Cancer: 1983. Pbk ISBN 92-9018-074-9. SwFr. 44, \$22.

Heterosis: Reappraisal of Theory and Practice. R. FRANKEL (ed.). Springer-Verlag: 1983. Pp.290. ISBN 3-540-12125-0/0-387-12125-0. DM 118, \$50.90.

Hypertensive Intracerebral Hemorrhage. M. MIZUKAMI et al. (eds). Raven: 1983. Pp.272. ISBN 0-89004-812-6. \$53.50.

Neuroendocrinology of Aging. J. MEITES (ed.). Plenum: 1983. Pp.383. ISBN 0-306-41310-8. \$45.

The Neurohypophysis: Structure, Function and Control. Progress in Brain Research, Vol.60. B.A. CROSS and G. LENG (eds). Elsevier Biomedical/North-Holland: 1983. Pp.542. ISBN 0-444-80479-X. \$119.25, Dfl.280.

Nutrition and Brain Development. By G.A. DHOSPESHWARKAR. Plenum: 1983. Pp.196. ISBN 0-306-41060-5. \$29.50.

Oxygen Transport to Tissue — IV. Advances in Experimental Medicine and Biology, Vol.157. H.I. BICHER and D.F. BRULEY (eds). Plenum: 1983. Pp.633. ISBN 0-306-41192-X. \$85.

Paleolimnology. Developments in Hydrobiology, Vol.15. Proceedings of the 3rd International Symposium held in Finland. J. MERILAINEN, P. HUTTUNEN and R.W. BATTARBEE (eds). Dr W. Junk: 1983. Pp.318. ISBN 90-6193-766-3. DG 200, \$80.

Parasitology: A Global Perspective. K.S. WARREN and J.Z. BOWERS (eds). Springer-Verlag: 1983. Pp.292. ISBN 0-387-90840-4/3-540-90840-4. DM 98, \$33.20.

Penguins: Past and Present, Here and There. By G.G. SIMPSON. Yale University Press: 1983. Pp.145. Hbk ISBN 0-300-01969-6; pbk ISBN 0-300-02095-9. Hbk np; pbk £5.95.

The Peripheral Arterial Chemoreceptors. Proceedings of the 7th International Symposium held September 1982, Leicester. D.J. PALLOT (ed.). Croom Helm/Oxford University Press: 1983. Pp.414. ISBN 0-7099-2041-5/0-19-520440-9. £27.50.

Physiological Plant Ecology IV. Ecosystem Process: Mineral Cycling, Productivity and Man's Influence. O.L. LANGE et al. (eds). Springer-Verlag: 1983. Pp.644. ISBN 3-540-10908-0/0-387-10908-0. DM 290, \$112.60.

Physiology of Adrenocortical Secretion. Frontiers of Hormone Research, Vol.11. By T. SUZUKI. Karger: 1983. Pp.216. ISBN 3-8055-3644-5. SwFr. 132, \$79.25, DM 158.

The Physiology of Excitable Cells. Proceedings of a Symposium held November 1982, Santa Monica. A.D. GRINNEL and W.J. MOODY Jr (eds). Alan R. Liss: 1983. Pp.620. ISBN 0-8451-2704-7. £55.

The Pineal Gland and its Endocrine Role. NATO Advanced Science Institutes Series, Vol.65. J. AXELROD, F. FRASCHINI and G.P. VELO (eds). Plenum: 1983. Pp.604. ISBN 0-306-41359-0. \$82.50.

Placental Physiology: Structure and Function of Fetomaternal Exchange. By J.J. FABER and K.L. THORNBERG. Raven: 1983. Pp.208. ISBN 0-89004-978-5. \$41.

Plant Peroxisomes. By A.H.C. HUANG, R.N. TRELEASE and T.S. MOORE Jr. Academic: 1983. Pp.252. ISBN 0-12-358260-1. \$34.50.

Prostaglandins and Related Substances. New Comprehensive Biochemistry, Vol.5. C. PACE-ASCIAC and E. GRANSTRÖM (eds). Elsevier Biomedical/North Holland: 1983. Pp.229. ISBN 0-444-80517-6. \$55, Dfl.129.

Receptor-Mediated Endocytosis. Receptors and Recognition Series B, Vol.15. P. CUATRECASAS and T. ROTH (eds). Chapman & Hall: 1983. Pp.304. ISBN 0-412-24820-4. £35.

Receptors as Supramolecular Entities. Advances in the Biosciences, Vol.44. Proceedings of the Biannual Capo Boi Conference held June 1981, Italy. G. BIGGIO et al. (eds). Pergamon: 1983. Pp.476. ISBN 0-08-029804-4. £55, \$100.

Reptilian Lungs: Functional Anatomy and Evolution. Advances in Anatomy, Embryology and Cell Biology, Vol.79. By S.F. PERRY. Springer-Verlag: 1983. Pp.81. Pbk ISBN 3-540-12194-3/0-387-12194-3. DM 50, \$21.60.

The Return of the Sea Eagle. By J.A. LOVE. Cambridge University Press: 1983. Pp.227. ISBN 0-521-25513-9. £15, \$29.95.

Sandy Beaches as Ecosystems. Developments in Hydrobiology, Vol.19. A. McLACHLAN and T. ERASMUS (eds). Dr W. Junk: 1983. Pp.757. ISBN 90-6193-770-1. DG 300, \$120.

Social Insects: Ecology and Behavioural Biology. By M.V. BRIAN. Chapman & Hall: 1983. Pp.377. Hbk ISBN 0-412-22920-X; pbk ISBN 0-412-22930-7. Hbk £25; pbk £12.95.

Solid Phase Biochemistry: Analytical and Synthetic Aspects. Chemical Analysis, Vol.66. W.H. SCOUTEN (ed.). Wiley: 1983. Pp.779. ISBN 0-471-08585-5. £71.25, \$99.75.

Somatosensory Integration in the Thalamus: A Re-evaluation Based on the New Methodological Approaches. G. MACCHI, A. RUSTIONI and R. SPREADICO (eds). Elsevier Biomedical/North-Holland: 1983. Pp.396. ISBN 0-444-80485-4. \$115, Dfl.270.

Symbiosis in Parent-Offspring Interactions. L.A. ROSENBLUM and H. MOLTZ (eds). Plenum: 1983. Pp.284. ISBN 0-306-41410-4. \$29.50.

Teratocarcinoma Stem Cells. Cold Spring Harbor Conferences on Cell Proliferation, Vol.10. L.M. SILVER, F.R. MARTIN and S. STRICKLAND (eds). Cold Spring Harbor Laboratory: 1983. Pp.743. ISBN 0-87969-160-3. \$85 (US), \$102 (elsewhere).

T Lymphocytes Today. J.R. INGLIS (ed.). Elsevier Biomedical/North Holland: 1983. Pp.193. Pbk ISBN 0-444-80525-9. \$19.50, £9.50.

Understanding Leukemia. By C.P. MARGOLIES and K.B. MCCREDIE. Charles Scribner's Sons: 1983. Pp.248. ISBN 0-684-17978-4. \$15.95.

Vegetationskunde. By K.H. KREEB. Verlag Eugen Ulmer GmbH: 1983. Pp.331. ISBN 3-8001-2499-8. DM 56.

Virology: An Information Profile. By R. NICHOLAS and D. NICHOLAS. Mansell: 1983. Pp.236. ISBN 0-7201-1673-2. £15.

Xylem Structure and the Ascent of Sap. By M.H. ZIMMERMANN. Springer-Verlag: 1983. Pp.143. ISBN 3-540-12268-0/0-387-12268-0. DM 54, \$22.30.

Yeasts: Characteristics and Identification. By J.A. BARNETT, R.W. PAYNE and D. YARROW. Cambridge University Press: 1983. Pp.811. ISBN 0-521-25296-2. £75, \$150.

Computer Science

Computational Linguistics. N. CERONE (ed.). Pergamon: 1983. Pp.245. ISBN 0-08-030253-X. £21, \$37.50.

The Computer Modelling of Mathematical Reasoning. By A. BUNDY. Academic: 1983. Pp.320. Hbk ISBN 0-12-141250-4; pbk ISBN 0-12-141352-0. Hbk np; pbk £9.50, \$15.

Real Time Programming 1981. Proceedings of the IFAC/IFIP Workshop held August-September 1981, Japan. Annual Review in Automatic Programming, Vol.11. T. HASEGAWA (ed.). Pergamon: 1983. Pp.138. ISBN 0-08-030411-7. £35, \$71.

General

North Sea Dynamics. By J. SUNDERMANN and W. LENZ. Springer-Verlag: 1983. Pp.693. ISBN 3-540-12013-0/0-387-12013-0. DM 98, \$42.30.

Ocean Science, 2nd Edn. By K. STONE. Wiley: 1983. Pp.673. ISBN 0-471-86719-5.

Reliability Theory and Its Application in Structural and Soil Mechanics. P. THOFT-CHRISTENSEN (ed.). Martinus Nijhoff: 1983. Pp.662. ISBN 90-247-2859-2. Dfl. 200, \$87.

Science and Public Policy. Annals of the New York Academy of Sciences, Vol.403. F.S. STERRETT (ed.). New York Academy of Sciences: 1983. Pp.120. Hbk ISBN 0-89766-198-2; pbk ISBN 0-89766-199-0. Hbk np; pbk \$24.

Soil Micromorphology, Vol.1: Techniques and Applications. P. BULLOCK and C.P. MURPHY (eds). A B Academic: 1983. Pp.366. ISBN 0-907360-06-8. £29.50.

Soil Micromorphology, Vol.2: Soil Genesis. P. BULLOCK and C.P. MURPHY (eds). A B Academic: 1983. Pp.705. ISBN 0-907360-07-6. £29.50.

New lab equipment in Munich

Analytica '84 is a major exhibition of laboratory equipment, to be held in Munich, West Germany, from 10 April to 13 April. Here we preview some of the products on view.

● Malvern Instruments' new 4600 molecular analyser combines in one instrument the two techniques of angular dependent total intensity light-scattering spectroscopy, and fluctuation spectroscopy (or photon correlation spectroscopy). Simultaneous measurements can be made of particle size (1–3,000 nm), molecular weight (10^2 – 10^{12}), diffusion coefficient (10^{-6} – 10^{-9}), and virial coefficient.

Circle No. 128 on Reader Service Card.

● Malthus Instruments show the latest version of their microbiological growth analyser, which incorporates a new "intelligent control system" to allow complete flexibility so that the system can be built to suit the size of the work load and expanded as this increases. This system allows one computer and microprocessor to control various water incubators running at different temperatures from 4°C to 45°C and allows varying scanning times to be carried out on cells of different volumes containing different samples. Results on any of the cells can be examined at any time and graphics can also be obtained on command. Malthus systems are available for blood culture, urine screening and antibiotic sensitivity testing, and more than 30 method sheets are available from the stand.

Circle No. 129 on Reader Service Card.

● As European distributors, Sterilin will be exhibiting a comprehensive range of Belco biotechnology products on its stand at Analytica '84. Equipment on view includes incubators, roller bottles, stirrers, plaque viewers, media bottles and a selection of flasks and special nose-piece pipettes. Sterilin products on display will include the full range of tissue culture plastics, the new ESR unit (designed to reduce the risk of leakage during transit and facilitate pipette insertion), the ELISA reader and examples of Esco tubing, pump tubing and general accessories.

Circle No. 130 on Reader Service Card.

● A sophisticated network of gas chromatography equipment has been announced by United Technologies Packard. The network is controlled by the Model 439GC or a stand-alone operating station that can be equipped for bidirectional transmission to an external computer. Up to 12 satellite gas chromatographs can be included in the network, and three satellite models for different analytical capabilities are available. The satellites can also be run without a network connection.

Circle No. 131 on Reader Service Card.

● Two new water-jacketed automatic CO₂ incubators have been introduced by Precision Scientific. The 1020 is a single-chamber version and the 2020 is a vertically stacked double-chamber version featuring independent controls for each separate chamber. The automatic CO₂ control system is solid-state push button digital set point and LED display of CO₂ tension to within $\pm 0.1\%$ with completely automatic CO₂ injection. An automatic air purge system prevents excessive CO₂ tension. CO₂ range is 0 to 20%. No external air supply is required, but it may be utilized if required. Temperature range is ambient $+5$ to 60°C controlled $+0.1^\circ$. Also from Precision Scientific is a new version of the DD-20 portable vacuum pump. This compact direct drive vacuum pump has free air capacity of 20 litres per minute and will reach an ultimate vacuum of 0.667 pascals. Precision Scientific products are shown at Analytica '84 on the Uniequip stand.

Circle No. 132 on Reader Service Card.

● The Shandon range for HPLC includes Hypersil column packing materials and pre-packed columns in 3 and 5 μ m sizes. Both pre-packed and empty columns are available with Shandon hand-tight and ordinary compression fittings. Also featured in the Analytica display, is a custom designed HPLC column packing pump. Amongst the electrophoresis and chromatography equipment on show, will be the Model 600 electrophoresis tank, available with a variety of custom designed accessories to perform virtually any flat-bed technique. The latest Vokam range for electrophoresis will be displayed and the stand will feature demonstrations of polyvalent counter immunoelectrophoresis pneumococcal separations.

Circle No. 133 on Reader Service Card.

● Mettler's AE166 is a new electronic analytical balance with a maximum capacity of 162 g, readable to 1 mg. The balance is fitted with DeltaRange, a fine range of 60 g, readable to 0.1 mg. When weighing-in, the display automatically switches over to the readability of the normal range when the fine range is exceeded. Formula weighing is especially easy because DeltaRange can be recalled by pressing the control bar after weighing-in each component. This allows the next component to be weighed in from zero with a readability of 0.1 mg, again up to 60 g. The AE166 DeltaRange is particularly useful when small amounts must be weighed into relatively heavy containers.

Circle No. 134 on Reader Service Card.

● Micro flow capability is now provided in the Spectra-Physics liquid chromatography (LC) product line. Micro liquid chromatography is becoming more important in industrial laboratories due to the cost savings and increased productivity. Micro capability is available as a retrofit kit for Spectra-Physics' SP8100 and SP8700 gradient liquid chromatographs. The kit extends the specified flow range of chromatographs down to 50 μ l per min, well below the normal practical flow rates for 2.1 mm columns. The kit was introduced at the Pittsburgh Conference in Atlantic City in early March.

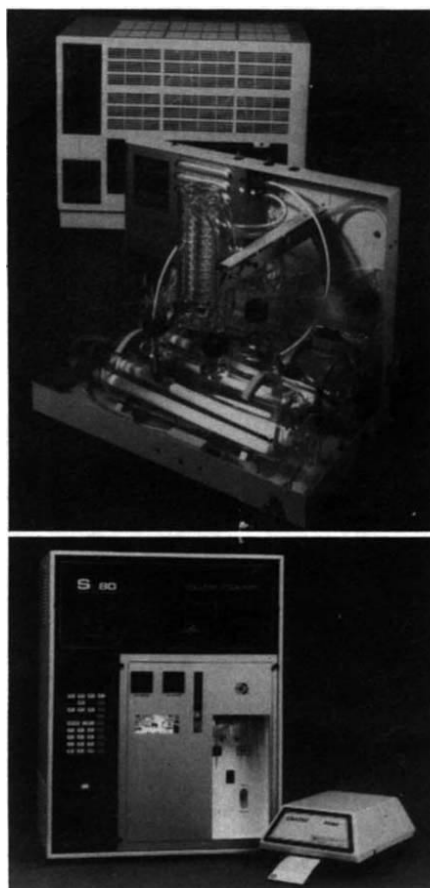
Circle No. 135 on Reader Service Card.

● *Analytical News* is published quarterly by Finnigan MAT. The publication is aimed at users of mass spectrometers and it contains articles on new products, seminars, applications, maintenance and operation and courses. The most recent issue, Winter 1983, features two articles on text editors and one on a computer program for real-time calculation of mass spectral abundances.

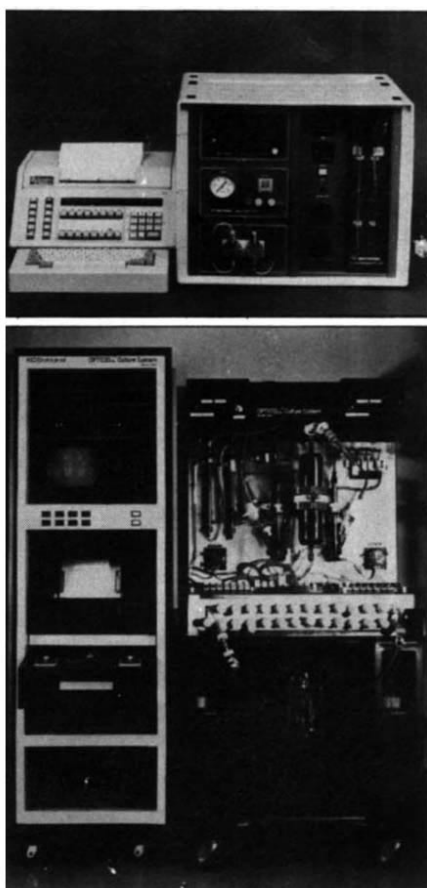
Circle No. 136 on Reader Service Card.



Two forms of spectroscopy combined — the Malvern 4600



The Jencons peristaltic pump and Biotronik's new ion chromatograph (top) and the S880 Coulter counter and the Opticell culture system.



● The Model S880 counter from Coulter Electronics offers the eight most requested parameters for the haematology laboratory, including whole blood platelet count, with the convenience of keypad control, and automated analysis from sample aspiration through to hard-copy print-out. The S880 is microprocessor-controlled via a simple to use keyboard and has an easy to read LCD display, allowing the operator to control and monitor the analyser's functions at all times. The Coulter S880 is compatible with the Accucomp system for full quality control assurance. Also new from Coulter, the EPICS 'C' is a low-cost, automatic cell analyser and sorter for clinical use. The EPICS 'C' analyses and sorts cells routinely on the basis of number, size and fluorescence, by the use of operator-programmable protocols. It reduces the reliance on fluorescent microscope analyses and the need for darkened rooms.

Circle No. 141 on Reader Service Card.

● Monoclonal antibodies are the basis of a number of new products from Becton Dickinson. A kit is now available for leukaemia immunophenotyping with anti-leukocyte surface antigen monoclonal antibodies — the pattern of reactivity of the kit's seven reagents is useful for immunophenotyping most human leukaemias. Another new reagent allows the determination of the ratio of T helper cells to T cytotoxic/suppressor cells on one slide. The single vial contains two monoclonal antibodies conjugated with different fluorochromes: Anti-Leu-2a FITC (green, T cytotoxic/suppressor cells) and Anti-Leu-3a phycoerythrin (yellow-orange, T helper cells). Samples can be examined under a standard fluorescence microscope.

Circle No. 142 on Reader Service Card.

● New products for large-scale cell culture, and serum-free media are to be found on the KC Biological stand. Opticell is a culture system designed for production of cells and cell products on a pilot scale. It is based on the Opticore growth chamber which is also available separately. Opticare is a sterile, ready to use chamber that very simply snaps into the closed loop. The ceramic substrate material, selected for its biocompatibility, is in the form of a cylinder with narrow channels running lengthwise through the cylinder. These channels are highly uniform and are in the shape of squares. Two models of Opticare are available: one with a surface area of 42,500 cm² (dimensions 9 cm diameter × 36 cm long), or with a surface area of 119,000 cm² (14.5 cm × 36 cm). KC Biological's Gelibead microcarrier is a new gelatin microcarrier offering a wide variety of cell propagation and harvesting advantages. One of the major features is in efficient cell harvesting — the Gelibead is dissolvable in a protease solution to facilitate cell removal and eliminates the need to separate cells from spent microcarriers.

Circle No. 143 on Reader Service Card.

● For applications where a compact ion chromatograph is needed — for instance in a mobile laboratory — Biotronik produces the IC 1000. Two basic systems are available: IC 1001 for the determination of organic acids, and IC 1002 for the determination of anions and cations. Other new items on the Biotronik stand include the BT7041 automatic sample injector for HPLC and a new amino acid analyser.

Circle No. 137 on Reader Service Card.

● Tubulazole-C, a new specific microtubule inhibitor from Janssen, inhibits the rate and extent of rat brain tubulin *in vitro* with an ID₅₀ of 3 × 10⁻⁷ M. In parallel experiments it is about 3 times as potent as nocodazole and 12 times as potent as colchicine. The compound induces secondary effects similar to those of classical microtubule inhibitors; loss of cell polarity, mitotic block and multinucleation, dispersal of Golgi elements and lysosomes, accumulation of intermediate filament coils and annulate lamellae. Also new from Janssen is a preparation of colloidal gold coated with staphylococcal protein A. The new preparation is of electron microscopy grade and suitable for a wide range of localization procedures.

Circle No. 138 on Reader Service Card.

These notes are based on information provided by the manufacturers. For further details circle the appropriate numbers on the Reader Service Card bound inside the journal.

● ASTEC, a manufacturer of a range of filtration frame cupboards, is showing the Astecair-5000 at Analytica. Astecair fume cupboards have no ducting, thereby eliminating the need for expensive installation work. The units are mobile, each carrying an integral lighting system. ASTEC produces 10 different filter media to cope with the harmful effects of solvents, acids, alkalis, ammonia, H₂S, SO₂, formaldehyde, resins, mercury vapours, radioactive iodine, aerosols, particles, bacteria, odours and low molecular weight fumes. The Astecair-5000 can be modified to include a HEPA filter, so preventing possible contamination in a clean area.

Circle No. 139 on Reader Service Card.

● The Jencons stand will show the "Sequential Perifill" peristaltic pump for use in laboratories where a dispenser is needed to repetitively dispense pre-set volumes. Also on show will be a range of versatile vac-seal glass Dewars for use with low-temperature liquids and gases, and the Zippette semi-automatic syringe mounted dispensers. The new Duo-Style Zippette offers alternative dispensing methods in the 5 ml, 10 ml, 30 ml and 50 ml models, with a choice of glass or PTFE piston. The Double Distillation Still gives 4 litres per hour of water twice distilled from glass and the Double D-Ionstill is designed for double distillation of deionized water and other forms of purified water.

Circle No. 140 on Reader Service Card.

● A new series of x-y recorders, LY171/181, has been introduced by Linseis of Frankfurt. They replace the series LY17 (A4) and LY18 (A3). Now even the A4 model LY171 is available with a paper wind-up mechanism. The slider is now made of an very stiff plastic material which is extremely light, providing a higher writing speed and acceleration. The pen-lift mechanism now has a stronger magnet making the recorder usable as a plotter.

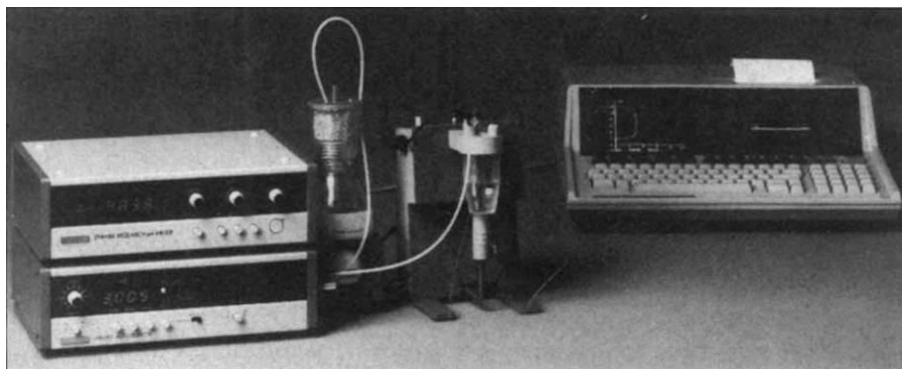
Circle No. 144 on Reader Service Card.

● Life Technologies is a new name at Analytica '84 and combines the cell culture, microbiology and molecular biology products of Gibco and Bethesda Research Laboratories. Two new fetal calf serum specifications have been introduced — Myoclon, selected for its ability to support myeloma cell lines, and aseptically collected FCS 309, a partially chemically defined serum low in endotoxins. For serum-free culture conditions, Select-a-Factor, containing 14 of the commonly incorporated factors, enables identification of those elements responsible for maximum growth of specific cell lines. For clinical microbiology, three new lyophilized supplements for the selective isolation of *Campylobacter* have been introduced. Life Technologies will also be exhibiting restriction enzymes, electrophoresis apparatus, cell culture reagents, microbial media and monoclonal antibodies.

Circle No. 145 on Reader Service Card.

● Radiometer's new D470 titration kit makes it possible to connect titration equipment to a Hewlett-Packard HP-85 personal computer. The kit includes all interface connections and fully documented software necessary to convert the HP-85 into a powerful titrator. The standard tape included with the kit is preprogrammed to perform automatic titrations in the simplest possible manner, as only four soft keys need to be manipulated in routine titrations. Results and curves are displayed on the screen of the HP-85 and can be printed by the built-in printer/plotter.

Circle No. 146 on Reader Service Card.



Radiometer's titration kit connects to the HP-85 personal computer.

● A computerized system for the identification of non-fermenting Gram-negative bacilli has been introduced by API. The API 20NE, designed to identify non-enteric Gram-negative bacilli, contains a database of 51 taxa, mainly non-fermenters but including a few fermentative organisms that are usually grouped alongside the non-fermenters in classification and identification schemes. 14 of the 20 tests on API 20NE were developed specifically for use with this group of organisms, relying largely on assimilation rather than fermentation of the test substrates.

Circle No. 147 on Reader Service Card.

● Samples of up to 2.5 ml in volume can be made protein-free and non-pyrogenous using the Sartorius Centrisart disposable ultrafiltration device. Centrisart consists of a centrifuge tube which holds the specimen and a filter unit (floater tube) which slides down inside it, coming to rest on the specimen. The device is then placed in a standard laboratory centrifuge which provides the driving force for filtration. The floater, with its membrane on the underside, is forced down through the specimen, the filtrate is collected inside the floater tube, while the concentrate remains in the outer tube. Components capable of settling are collected on the bottom of the centrifuge tube without coming into contact with the membrane. Centrisart will also concentrate protein solutions.

Circle No. 148 on Reader Service Card.

● The Futura range from Beckman is a family of eleven pH electrodes, including glass, reference and combination electrodes for a wide variety of applications. One electrode is specially designed to measure samples of low ionic strength, such as pure water and acid rain.

Circle No. 149 on Reader Service Card.

● A new automatic liquid injection system has been developed by Carlo Erba for its high resolution gas chromatographs — the Autosampler AS 550. The system allows unattended cold on-column injections of up to 60 samples.

Circle No. 150 on Reader Service Card.

ADVERTISEMENTS

MICROPARTICLE Technology Limited

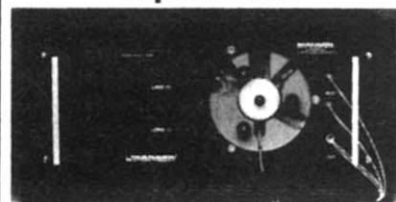
Monodisperse microparticles in a diverse range of sizes and materials. Monodisperse polystyrene particles (Coefficient of variation $\leq 10\%$) in range 3μ to 85μ diameter. Other sizes and materials available on request. Also, fluorescent and dyed particles available. All competitively priced.

Send for our brochure or telephone direct with your order.

031-557 3796

Unit 13, Canongate Venture, New St., Edinburgh EH8 8EH.

Liposomes



LIPOPREP the equipment for the reproducible preparation of stable unilamellar liposomes of defined size in the range of 30-300 nm ϕ with extremely narrow particle size distribution for e.g. the efficient incorporation of (therapeutic) agents incl. antibodies.

DIANORM-Geräte, P.O. Box 126, D-8 München 65, FRG

PHORBOL ESTERS

99 + % PURE
MPMA

4-O-METHYLPHORBOL 12-MYRISTATE
13-ACETATE

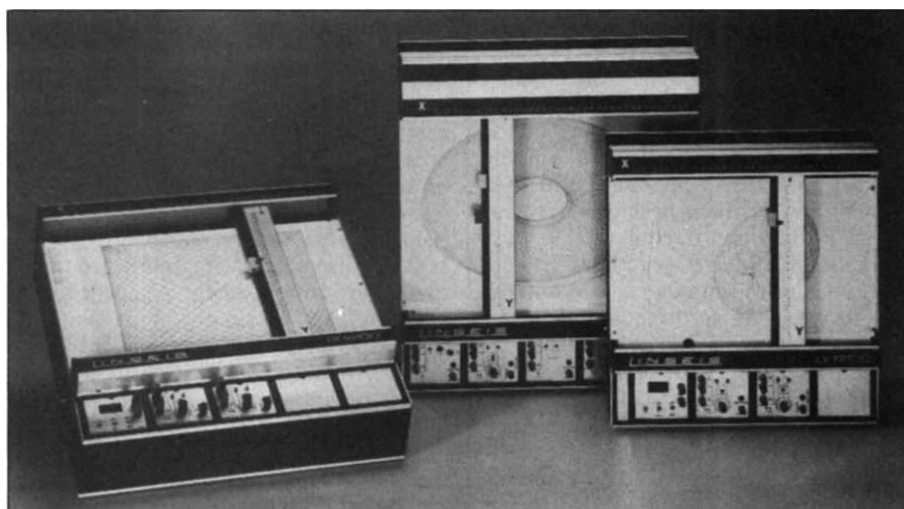
Cat. #M2550 5mg/565^{mm}



LC SERVICES CORPORATION

165 New Boston Street
Woburn, MA 01801
(617) 938-1700

Circle No. 47 on Reader Service Card.



The LY171/181 range of recorders.

An international bio-labour market

Richard Pearson*

The rapid development of biotechnology is creating a demand for new skills. But the industry matures the emphasis will move towards more traditional skills in bioprocess engineering.

SCIENCE has always been an international activity, with a regular interchange of staff and ideas between countries. In most subjects, however, such job mobility is restricted to a limited number of senior academics, together with a larger flow of post-docs seeking training places, chiefly in the United States. By 1981, more than half the doctorates awarded in engineering in the United States went to foreigners and almost 60 per cent of the postdoctoral students were from overseas. Indeed, it has been estimated that more US Government grants are given to foreign students to study in the United States than to Americans to study overseas. On a smaller scale, the United Kingdom has, for many years, acted as a centre for postgraduate and postdoctoral training, especially for students from the Commonwealth. This interchange of staff is seen to be beneficial to all concerned, providing both training and research assistance, and a cross fertilization of ideas and experience between countries.

Occasionally, however, a sudden surge in the commercialization of a particular area of science, or increased allocation of funds in one country, has caused imbalances in the supply and demand for key skills. As a result, countries with a competitive advantage in the labour market, most notably in North America, can trigger off a "brain drain" from less advantaged labour markets. Thus, in the past thirty years, the United Kingdom has seen the alarm bells ringing over the brain drain of its doctors, nuclear scientists, electronics specialists and, most recently, biotechnologists. West Germany has joined Britain in expressing concern, while France, with a much smaller pool of indigenous skills, is actively encouraging its own nationals working overseas with biotechnology skills to return home.

The latest report from the US Office of Technology Assessment (OTA), *Commercial Biotechnology* (see *Nature* 2 February, p.399), provides a valuable international perspective on the rapidly developing world labour market for biotechnology skills. Biotechnology is of course an interdisciplinary activity, embracing diverse skills in biochemistry, biology, microbiology, genetics, medical and veterinary sciences and bioprocess engineering. The scientific side, in

particular, requires high qualifications; it is unlikely that anybody without at least three years postdoctoral experience would be considered as a specialist in biotechnology. Despite its "newness" and classification problems, some estimates are beginning to emerge of the numbers involved at a professional level. The OTA study estimates that there are about 5,000 people employed by companies in the United States in biotechnology research and development. There are no estimates, however, for the academic world nor the public laboratories. OTA was unable to find any estimates for France, Germany, Switzerland, Japan or the United Kingdom, the main countries with a significant involvement in biotechnology. A report soon to be published by the UK Science and Engineering Research Council puts the total number of professionals in all sectors at just under 2,000, with nearly half in commercial companies. A crude estimate therefore suggests that the total number worldwide may be under 30,000, a smaller figure than for those working in semiconductors, perhaps the most closely analogous sector.

The OTA report provides a useful profile of the skills of the 5,000 people in commercial research and development. It shows just over a third with skills in genetic manipulation, and a third related to processing activities. Much of the research and public attention so far has been focused on the advances in genetic manipulation, and the new venture "glamour" companies such as Biogen, Celltech and Cetus have been able to offer highly attractive and well publicized terms of employment, often including equity shares, to attract the best genetics specialists from all corners of the world. Salary levels of up to \$100,000 a year have been quoted. At the same time, many of the established industrial companies, notably those in pharmaceuticals, have also been investing and recruiting heavily in this area, both nationally and internationally. Such international mobility, particularly where the flow is not reciprocated, will only be sustained as long as there is a clear imbalance in the skills available, allied with differential reward structures between the countries. These rewards of course need not be just monetary; they also include quality of life and environment, research facilities and the freedom to develop new ideas and take risks.

Looking internationally, the OTA

report says that the United States has the largest number of genetics specialists, and after a period of shortages few companies are now reporting shortages of staff with the basic skills. Similarly, the United Kingdom and West Germany also seem to have an adequate pool of such skills. The one major exception is said to be Japan, where an earlier lack of investment in basic science did not allow a basic pool of skills to develop. To remedy the situation, the Japanese have started in-house training, sending people abroad to be trained, including five corporate researchers sent to Genex for a three-month course at a cost of \$120,000 each. Mobility between companies is also being encouraged, a rare phenomenon in Japan. No attempt has so far been made to attract foreign nationals, although individual foreign consultants, working in their home countries, are being engaged.

It is in the bioprocess engineering area, the application of biotechnology, that future international shortages seem most likely to occur. In the United States, it is thought that shortages of these skills may cause a bottleneck to the rapid commercialization of biotechnology and several US companies are actively recruiting in Europe to overcome such shortfalls. The United Kingdom seems also to have a potential shortage of such skills, as does France, which has an all-round skill shortage, but West Germany has a more adequate supply. Ironically Japan, which lacks the basic skills, is, as in other technologies, well supplied with the applied staff, in this case bioprocess engineers, many of whom have a background in microbial physiology, an area of potential skill shortage in most Western countries. As bioprocess engineering is still in its infancy, and seems to require more practical experience than the more "academic" genetic manipulation areas, it seems that there will be a premium on anyone with such skills in the future and international mobility could well increase.

In its short history, biotechnology has seen one international migration prompted by the rush of investment in genetic manipulation. As the supply of people with these skills has increased to near-sufficiency in many countries, the flow has abated. The market may have gone quiet for now, but as biotechnology matures, a second international migration could take place as the leading companies seek to recruit in the limited world pool of experienced bioprocess engineers. □

* Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.